

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 10:55:05 ; Search time 3842.15 Seconds  
(without alignments)  
1906.419 Million cell updates/sec

Title: US-09-394-745-6154  
Perfect score: 444  
Sequence: 1 cgaaaacactggtacccaaa.....tcccattttaagaaataaat 444

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_om:\*  
20: em\_or:\*  
21: em\_ov:\*  
22: em\_pat:\*  
23: em\_ph:\*  
24: em\_pl:\*  
25: em\_ro:\*  
26: em\_sts:\*  
27: em\_sy:\*

28: em\_un:\*  
 29: em\_vi:\*  
 30: em\_htgo\_hum:\*  
 31: em\_htgo\_inv:\*  
 32: em\_htgo\_rod:\*  
 33: em\_htg\_hum:\*  
 34: em\_htg\_inv:\*  
 35: em\_htg\_rod:\*  
 36: em\_htg\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result	%		Query			ID	Description
	No.	Score	Match	Length	DB		
	1	353.4	79.6	481	8	ZMA133529	AJ133529 Zea mays
	2	284	64.0	379	6	AX015683	AX015683 Sequence
	3	58.8	13.2	592	8	ZMA297902	AJ297902 Zea mays
	4	42.4	9.5	579	8	ZMA297903	AJ297903 Zea mays
c	5	41.4	9.3	7218	6	I66494	I66494 Sequence 14
c	6	39.4	8.9	41334	3	CELC30B5	U23450 Caenorhabdi
	7	37.8	8.5	188254	2	AC080021	AC080021 Mus muscu
	8	36.8	8.3	170225	2	AC020602	AC020602 Homo sapi
	9	36	8.1	49307	3	CELY34D9A	AC024756 Caenorhab
c	10	36	8.1	166214	2	AC006735	AC006735 Caenorhab
	11	36	8.1	183980	2	AC011081	AC011081 Homo sapi
c	12	35.6	8.0	63325	9	AL353592	AL353592 Human DNA
	13	35.4	8.0	563	8	ZMA297901	AJ297901 Zea mays
	14	35.4	8.0	78874	2	AL355521	AL355521 Homo sapi
	15	35.4	8.0	174986	2	AC064821	AC064821 Homo sapi
	16	35.4	8.0	175352	2	AC092491	AC092491 Homo sapi
	17	35.2	7.9	1680	8	AB008680	AB008680 Glycine m
	18	35.2	7.9	3636	8	SOYBPSP	M13759 Glycine max
c	19	35	7.9	105383	2	AC010057	AC010057 Drosophil
	20	35	7.9	145087	2	AC019753	AC019753 Drosophil
c	21	35	7.9	159108	2	AC026031	AC026031 Homo sapi
c	22	35	7.9	168047	3	AC091219	AC091219 Drosophil
c	23	35	7.9	302527	3	AE003469	AE003469 Drosophil
c	24	34.8	7.8	112203	9	HSJ519P24	AL050401 Human DNA
	25	34.6	7.8	1141	6	AX083744	AX083744 Sequence
	26	34.6	7.8	161799	9	AC002091	AC002091 Genomic s
c	27	34.6	7.8	182341	2	AC073337	AC073337 Homo sapi
c	28	34.6	7.8	186552	2	AC090610	AC090610 Homo sapi
	29	34.6	7.8	191686	9	AL359197	AL359197 Human DNA
	30	34.4	7.7	89779	8	AB005234	AB005234 Arabidops
c	31	34.4	7.7	111547	2	AP002332	AP002332 Homo sapi
	32	34.4	7.7	169230	2	AC012211	AC012211 Homo sapi
	33	34.4	7.7	180041	2	AC009831	AC009831 Homo sapi
	34	34.4	7.7	189181	2	AP001333	AP001333 Homo sapi
c	35	34.2	7.7	1404	6	E03536	E03536 DNA sequenc
c	36	34.2	7.7	1404	6	E08057	E08057 DNA encodin
c	37	34.2	7.7	1404	6	E08058	E08058 DNA encodin
c	38	34.2	7.7	1404	6	E08059	E08059 DNA encodin

c	39	34.2	7.7	1404	6	E08060	E08060 DNA encodin
c	40	34.2	7.7	1404	6	I23831	I23831 Sequence 3
c	41	34.2	7.7	1404	6	I23832	I23832 Sequence 5
c	42	34.2	7.7	1404	6	I23833	I23833 Sequence 7
c	43	34.2	7.7	1404	6	I43342	I43342 Sequence 2
c	44	34.2	7.7	1404	6	I43343	I43343 Sequence 3
c	45	34.2	7.7	1404	6	I43344	I43344 Sequence 4

# ALIGNMENTS

RESULT 1  
ZMA133529

LOCUS ZMA133529 481 bp mRNA PLN 02-DEC-1999

DEFINITION Zea mays mRNA for BETL2 protein.

ACCESSION AJ133529

VERSION AJ133529.1 GI:5042328

KEYWORDS betl2 gene; BETL2 protein.

SOURCE Zea mays.

ORGANISM Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 481)  
AUTHORS Hueros,G., Royo,J., Maitz,M., Salamini,F. and Thompson,R.D.  
TITLE Evidence for factors regulating transfer cell-specific expression  
in maize endosperm  
JOURNAL Plant Mol. Biol. 41 (3), 403-414 (1999)  
MEDLINE 20064976

REFERENCE 2 (bases 1 to 481)  
AUTHORS Thompson,R.D.  
TITLE Direct Submission  
JOURNAL Submitted (08-MAR-1999) Thompson R.D., Plant Breeding, MPI For  
Plant Breeding Research, Carl-von-Linne-weg 10, D-50829 Koeln,  
GERMANY

FEATURES Location/Qualifiers

source 1..481  
/organism="Zea mays"  
/variety="A69Y"  
/db\_xref="taxon:4577"  
/country="Argentina"

sig\_peptide 44..121  
/gene="betl2"

CDS 44..331  
/gene="betl2"  
/codon\_start=1  
/product="BETL2 protein"  
/protein\_id="CAB44662.1"  
/db\_xref="GI:5042329"  
/db\_xref="SPTREMBL:Q9XGE0"  
/translation="MAKCSSFQGLFWLLSMILLASFVAHARTTSGQTKEDSNARNMTM  
TKTRASGNILVSRNDDGPCYLD SGLNEYVCRKTNKCYKSLVLCVASCQPSS"

gene 44..331  
/gene="betl2"

mat\_peptide 119..328  
/gene="betl2"

/product="BETL2 protein"  
 BASE COUNT      156 a      100 c      107 g      118 t  
 ORIGIN

Query Match                      79.6%;    Score 353.4;    DB 8;    Length 481;  
 Best Local Similarity    88.7%;    Pred. No. 4.2e-98;  
 Matches    392;    Conservative    0;    Mismatches    49;    Indels    1;    Gaps    1;

```

Qy      3 aaaacactggtacccaaaacaaccgtcaaccaagggcaaattcaacaacctccaaagaat 62
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     13 AGACTATTGTAGCTCATATCATCTGTCACCCATGGCGAAGTGCAGCAGCTTCCAAGGATT 72

Qy     63 aatccgggtgccttccaagaatcctccaaccacccttggtgcccaagcaagccacaaaaa 122
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     73 ATTCTGGTTGCTTTCCATGATTCTTCTAGCATCCTTTGTTGCTCATGCACG-CACAACAA 131

Qy    123 gtgggcaaaccaaagaggacagcaatgctaggaaaatgacgatgacaaagacgagggcat 182
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    132 GTGGGCAAACCAAAGAGGACAGCAATGCTAGGAACATGACGATGACCAAGACGAGGGCAT 191

Qy    183 cgggcaacataacttggttagccgtaatgacgacggggccatgctatctagattccgggtctta 242
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    192 CAGGCAACATACTTGTTAGCCGTAATGACGACGGGCCATGCTATCTAGATTCCGGTCTTA 251

Qy    243 atgagtacgtctgcagaaagactaataagtgcataagagcttggtgctctgcgtggcga 302
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    252 ATGAGTACGTCTGCAGAAAGACTAATAAGTGCTATAAGAGCTTGGTGCTCTGCGTGCGCA 311

Qy    303 gttgtcaaccatcatcatgaattcaagatactgcggagacatcatgatactgcggagaca 362
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    312 GTTGTCAACCATCATCATGAATTCATGATACTGCGGAGACATCATGATACTGCGGAGACA 371

Qy    363 gacggccagagatgangctagctagatgccgtttcaccannatattatgtaacacccaaa 422
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    372 GACGGCGAGAGATGAGGCTAGCTAGATGCTGTTTCACCAAATATTATGTAACACCCAAA 431

Qy    423 tctcccattttaagaaataaat 444
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    432 TCTCCCATTTTAAGAAATAAAT 453
  
```

RESULT    2  
 AX015683  
 LOCUS      AX015683      379 bp      DNA                      PAT              07-SEP-2000  
 DEFINITION    Sequence 1 from Patent WO9950427.  
 ACCESSION    AX015683  
 VERSION      AX015683.1    GI:10041512  
 KEYWORDS     .  
 SOURCE       Zea mays.  
              ORGANISM    Zea mays  
                           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
                           Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
                           clade; Panicoideae; Andropogoneae; Zea.  
 REFERENCE    1    (bases 1 to 379)  
 AUTHORS      Yan,G., Salamini,F., Thompson,R.D. and Hueros,G.  
 TITLE        Novel basal endosperm transfer cell layer (betl) specific genes

```

JOURNAL      Patent: WO 9950427-A 1 07-OCT-1999;
              YAN GUO (DE); MAX PLACK GES ZUR FOERDERUNG D (DE); SALAMINI
              FRANCESCO (DE); THOMPSON RICHARD D (DE); HUEROS GREGORIO (ES)
FEATURES
  source      Location/Qualifiers
              1. .379
              /organism="Zea mays"
              /db_xref="taxon:4577"
  CDS         44. .331
              /note="unnamed protein product"
              /codon_start=1
              /protein_id="CAC07599.1"
              /db_xref="GI:10041513"
              /translation="MAKCSSFQGLFWLLSMILLASFVAHARTTSGQTKEDSNARNMTM
              TKTRASGNILVSRNDDGPCYLD SGLNEYVCRKTNKCYKSLVLCVASCQPSS"
BASE COUNT   107 a      85 c      94 g      93 t
ORIGIN

```

```

RESULT      3
ZMA297902
LOCUS       ZMA297902      592 bp      mRNA                      PLN      11-JAN-2001
DEFINITION  Zea mays mRNA for basal layer antifungal peptide (bap-3a gene).
ACCESSION  AJ297902
VERSION    AJ297902.1   GI:12214248

```

```

KEYWORDS      bap-3a gene; basal layer antifungal peptide.
SOURCE        Zea mays.
ORGANISM      Zea mays
              Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
              clade; Panicoideae; Andropogoneae; Zea.
REFERENCE     1 (bases 1 to 592)
AUTHORS       Serna Sanz, A. and Thompson, R.D.
TITLE         Maize endosperm secretes a novel antifungal protein into adjacent
              maternal tissue
JOURNAL       Unpublished
REFERENCE     2 (bases 1 to 592)
AUTHORS       Serna, A.
TITLE         Direct Submission
JOURNAL       Submitted (31-OCT-2000) Serna A., Plant Physiology, Max Planck
              Institut, Carl von Linne Weg 10, Cologne, 50829, GERMANY
FEATURES      Location/Qualifiers
    source     1..592
               /organism="Zea mays"
               /variety="A69Y"
               /db_xref="taxon:4577"
               /tissue_type="endosperm"
               /dev_stage="7 days after pollination"
    sig_peptide       49..132
               /gene="bap-3a"
    CDS          49..339
               /gene="bap-3a"
               /function="putative antifungal peptide"
               /codon_start=1
               /product="basal layer antifungal peptide"
               /protein_id="CAC21606.1"
               /db_xref="GI:12214249"
               /translation="MVKILDHISIRGFFLLFMVLVASFVGHAQIIRGETKEDNDTKSM
               TMTTMRPGSYVTSMDKSSLCFEDIKTLWYICRTTYHLYRTLKDCLSHCNM"
    gene         49..339
               /gene="bap-3a"
    mat_peptide   133..336
               /gene="bap-3a"
               /product="bap-3a protein"
BASE COUNT     196 a      102 c      115 g      179 t
ORIGIN

```

Qy 93 cacccttggtgcccgaagcaagccacaaaaagtgggcaaaccaaagaggacagcaatgcta 152  
 || ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 Db 113 CATCCTTTGTTGGTCATGCACAGATAATAAGAGGTGAAACCAAAGAGGACAACGACACCA 172  
 Qy 153 ggaaaatgacgatgacaaagacgagggcatcgggcaacatacttgttagccgtaatgacg 212  
 || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
 Db 173 AGAGCATGACGATGACAACAATGAGACCAGGAAGCTATGTAACTAGCATGGATGAAAAAT 232  
 Qy 213 acgggccatgctatctagattccggtcttaatgagtacgtctgcagaaagactaataagt 272  
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 233 CTAGCTTGCTTTGAGGATATAAAAACTTTATGGTACATCTGCAGAACAACTTATCACC 292

Qy 273 gctataagagcttggtgctctgcgtggcgagttg 306  
 |||| || ||| | ||| || || |||

Db 293 TTTATAGGACATTGAAGGATTGCCTGTCGCATTG 326

RESULT 4  
 ZMA297903  
 LOCUS ZMA297903 579 bp mRNA PLN 11-JAN-2001  
 DEFINITION Zea mays mRNA for basal layer antifungal peptide (bap-3b gene).  
 ACCESSION AJ297903  
 VERSION AJ297903.1 GI:12214250  
 KEYWORDS bap-3b gene; basal layer antifungal peptide.  
 SOURCE Zea mays.  
 ORGANISM Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
 clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 579)  
 AUTHORS Serna Sanz,A. and Thompson,R.D.  
 TITLE Maize endosperm secretes a novel antifungal protein into adjacent  
 maternal tissue  
 JOURNAL Unpublished

REFERENCE 2 (bases 1 to 579)  
 AUTHORS Serna,A.  
 TITLE Direct Submission  
 JOURNAL Submitted (31-OCT-2000) Serna A., Plant Physiology, Max Planck  
 Institut, Carl von Linne Weg 10, Cologne, 50829, GERMANY

FEATURES Location/Qualifiers  
 source 1. .579  
 /organism="Zea mays"  
 /variety="A69Y"  
 /db\_xref="taxon:4577"  
 /tissue\_type="endosperm"  
 /dev\_stage="7 days after pollination"  
 sig\_peptide 25. .108  
 /gene="bap-3b"  
 CDS 25. .312  
 /gene="bap-3b"  
 /function="putative antifungal peptide"  
 /codon\_start=1  
 /product="basal layer antifungal peptide"  
 /protein\_id="CAC21607.1"  
 /db\_xref="GI:12214251"  
 /translation="MVKSLDHITIRGLFLLFMFLVASFVGHAQIIRGETKENKDTNSM  
 TMTTRPGSYVISMDEKSSLCFLDPRTLWYICKITYRLFRTLKDCLEFCHSI"  
 gene 25. .312  
 /gene="bap-3b"  
 mat\_peptide 109. .309  
 /gene="bap-3b"  
 /product="bap-3b protein"

BASE COUNT 189 a 99 c 114 g 177 t  
 ORIGIN

Query Match 9.5%; Score 42.4; DB 8; Length 579;

Best Local Similarity 52.2%; Pred. No. 0.035;  
Matches 119; Conservative 0; Mismatches 106; Indels 3; Gaps 1;

Qy	96	ccttggtgcccgaagcaagccacaaaaagtgggcaaaccgaagaggacagcaatgctagga	155
Db	92	CCTTTGTTGGTCATGCACAGATAATAAGAGGTGAAACCAAGGAGAATAAGGACACTAACA	151
Qy	156	aaatgacgatgacaaagacgaggggcatcgggcaacatacttgttagccgtaatgacgacg	215
Db	152	GCATGACGATGACA---ACAAGACCAGGAAGCTATGTAATTAGCATGGATGAAAAATCTA	208
Qy	216	ggccatgctatctagattccggtcttaatgagtacgtctgcagaaagactaataagtgct	275
Db	209	GCTTGTGCTTTCTGGATCCAAGAACTCTATGGTACATCTGCAAATAACATATCGCCTTT	268
Qy	276	ataagagcttggtgctctgctgctggcgagttgtcaaccatcatcatgaa	323
Db	269	TTAGGACATTGAAGGATTGCTTGGAGTTTTGCCACAGTATATGATGCA	316

```

RESULT      5
I66494/c
LOCUS       I66494      7218 bp      DNA
DEFINITION  Sequence 14 from patent US 5670367.
ACCESSION   I66494
VERSION     I66494.1  GI:2724471
KEYWORDS     .
SOURCE      Unknown.
  ORGANISM   Unknown.
             Unclassified.
REFERENCE   1  (bases 1 to 7218)
  AUTHORS   Dorner,F., Scheiflinger,F. and Falkner,F.Gunter.
  TITLE     Recombinant fowlpox virus
  JOURNAL   Patent: US 5670367-A 14 23-SEP-1997;
FEATURES             Location/Qualifiers
  source             1. .7218
                    /organism="unknown"
BASE COUNT      1944 a      1491 c      1486 g      1929 t      368 others
ORIGIN

```

Query Match 9.3%; Score 41.4; DB 6; Length 7218;  
Best Local Similarity 3.1%; Pred. No. 0.096;  
Matches 12; Conservative 211; Mismatches 163; Indels 0; Gaps 0;

[illegible]



[illegible]

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

WARNING: These data have only had automated annotation and have not yet been subjected to manual review of that annotation. We will be manually reviewing this information as quickly as possible and at that time this GenBank record will be updated and this warning removed.

NOTES:

Coding sequences below are predicted from computer analysis, using the program Genefinder(P. Green and L. Hillier, ms in preparation).

FEATURES	Location/Qualifiers
source	1. .41334 /organism="Caenorhabditis elegans" /strain="Bristol N2" /db_xref="taxon:6239" /chromosome="II" /clone="CEL30B5"
gene	2450. .3867 /gene="C30B5.3"
CDS	join(2450. .2484,2567. .2802,2852. .3627,3679. .3867) /gene="C30B5.3" /note="similar to C. elegans protein C40H1.1 (similar to ovarian protein (fly)); coded for by C. elegans cDNA yk302h1.5" /codon_start=1 /protein_id="AAK31467.1" /db_xref="GI:13592368" /translation="MHKDAAENYDKQLELRSSPQINQILRCTSNTNATSSEIQLRNRO AVIVSNFREPDRRLGRYSKYYYHHNVGPEVYSRKVFVGGLPSCVKESDILNFFSRYGR LQVDWPSKHYECKSDSDPSLCNEPISSSSYQPSHLLAMVSPPFGEINPFMRNMPAQSE SSQTGGFGRISGSGSIGGFLNPGMAQVARGNLGFGSTKSDGSINGDKRQHHLGYVLLF EKERSVRDLVLDCEEEEEGLFITLESSTDSIRVQIRPWLLADAEFLMDFNVPINTKLV AFIGGVPRPLKAVELAHFFEQTYGHVVCVGIDIDNKFYPRGSGRVAFSDDYDAYVQAI TDRYIVLDHEDIHKRVEIKPYFFHNQSCEECSGRYHRQHAPYFCPSLECFQYYCEPCW HKMHSHPSRFHHMPVVKGV"
gene	complement(4589. .5670) /gene="C30B5.4"
CDS	complement(join(4589. .5188,5315. .5444,5492. .5670)) /gene="C30B5.4" /note="Contains similarity to Pfam domain: PF00076 (rrm), Score=81.4, E-value=6.1e-21, N=1" /codon_start=1 /evidence=not_experimental /protein_id="AAK31468.1" /db_xref="GI:13592369" /translation="MNPITNIKNQRMNERELSLGYAGDLKKSWHQTYKDSAWIYIGG LSYALSEGDVIAVFSQYGEVMNINLIRDKDTGSKSGFAFLCYKDQRSTILAVDNFNGI SLHKRMIRVDHVEEYKVPKYKEDADDETKRLWEEGCAPKPMREAAAPMEVQEQRKKA KEVLLDIGDVDEELLKKIKKDKKKAKKEKKREKKRAKKIRKLEKKAARDPDGDWNNKA"

```

KLIDKVVAEDDLYGKHFDFGKKKEVEEVKHNPRPDFEKADWRDIEIWKVIREREKA
EKAARGETSEAWGPEDHYVSKRYQGR"
gene      5970. .6878
          /gene="C30B5.2"
CDS       join(5970. .5985,6152. .6227,6271. .6454,6541. .6579,
          6605. .6703,6762. .6878)
          /gene="C30B5.2"
          /note="coded for by C. elegans cDNA yk402a11.3; coded for
          by C. elegans cDNA yk402a11.5; coded for by C. elegans
          cDNA yk455a11.5; coded for by C. elegans cDNA yk745c8.5;
          coded for by C. elegans cDNA yk745c8.3"
          /codon_start=1
          /product="Hypothetical protein C30B5.2"
          /protein_id="AAK31472.1"
          /db_xref="GI:13592373"
          /translation="MGGVRAVAALAFAGVVGLTFLVLGCALPRYGTWTPMFVITFYVL
          SPVPLLIARRFQEDMTGTNACIELALFITTGIVISAFALPIVLAHAGTIANSACFLVN
          TGSHTCIVMTTVEAGASLCSKSTSLISCQVIVIAMSAFLIFIANSSINFSVIIF
          YFRIFNGEDMNGMSLW"
gene      7525. .10141
          /gene="C30B5.1"
CDS       join(7525. .7627,7678. .7766,7812. .7961,8028. .8289,
          8338. .8722,9104. .9364,9412. .9970,10022. .10141)
          /gene="C30B5.1"
          /note="coded for by C. elegans cDNA yk12e2.3; coded for by
          C. elegans cDNA yk12e2.5; coded for by C. elegans cDNA
          yk18d7.5; coded for by C. elegans cDNA yk44h5.3; coded for
          by C. elegans cDNA yk44h5.5; coded for by C. elegans cDNA
          yk54d12.5; coded for by C. elegans cDNA yk122g2.5; coded
          for by C. elegans cDNA yk146c11.5; coded for by C. elegans
          cDNA yk209b3.3; coded for by C. elegans cDNA yk209b3.5;
          coded for by C. elegans cDNA yk218c3.3; coded for by C.
          elegans cDNA yk218c3.5; coded for by C. elegans cDNA
          yk285h2.3; coded for by C. elegans cDNA yk285h10.5; coded
          for by C. elegans cDNA yk294c3.5; coded for by C. elegans
          cDNA yk296b11.3; coded for by C. elegans cDNA yk296b11.5;
          coded for by C. elegans cDNA yk297d3.5; coded for by C.
          elegans cDNA yk320f12.5; coded for by C. elegans cDNA
          yk346h3.3; coded for by C. elegans cDNA yk346h3.5; coded
          for by C. elegans cDNA yk347a4.5; coded for by C. elegans
          cDNA yk350a7.5; coded for by C. elegans cDNA yk351b1.5;
          coded for by C. elegans cDNA yk397e7.3; coded for by C.
          elegans cDNA yk397e7.5; coded for by C. elegans cDNA
          yk408e10.5; coded for by C. elegans cDNA yk419c3.3; coded
          for by C. elegans cDNA yk419c3.5; coded for by C. elegans
          cDNA yk430g3.3; coded for by C. elegans cDNA yk430g3.5;
          coded for by C. elegans cDNA yk442h4.5; coded for by C.
          elegans cDNA yk451h6.3; coded for by C. elegans cDNA
          yk451h6.5"
          /codon_start=1
          /product="Hypothetical protein C30B5.1"
          /protein_id="AAK31466.1"
          /db_xref="GI:13592367"
          /translation="MNKDNSPSESLVNSSASEDKTVKTSMSANNPSINDSSLPERLPA
          MCGCDSQNREESANSQVLSHDHANESLSKQADDVIYEGGLEDKDSVFRAIIEDDDACE
          VLDDGIDNCEVEIPVAIECHLCNEMMNLCRLRTRYRGQAREYPAYRCNRKGCQTFRSI
          RKVFGNCMSDNDNSGDATTRTMIYHPHVKVRKTSPEGEFSDDDSKEFVVQRVYIPKIDN

```

```

RNKDQSVKSLSVTDRMRKANQDRAIVFSEFADQLRRDIVANKRVRIKKHVEEDQQGTL
FYISKELNPQEVIELQDVIIKTLISMRKIPPPMTMDDLPLFANCPYSKKAVDAGEFDM
SETEERPRPKLPKTIYQNTRTHSDSLTERS RHFTYDELKHSMEANAEGPSTCRIGISR
SERNSESLQWIGLKTYEDSTDGFAKSVNSEGERQQGINGQIDDEVFEDES RFKPVASM
QDVMRTLSTPSTLRPPVSPLMKSFSSMDSSNFKNPINAHPMMFFSDPHFLINNNQFNS
PTPFYALNSYFSPFMDSSQVYQNATDEHFTPSGASEVVCTEQTPGHHTYIHPDSGV
RLTDPSAFLTSFVPSPI TKVPPINFNDETSRSRASAPSTSVYNTGEFERSHKKYENE
EDEKLMDNQNFSDQPKKE"
gene      complement(10349. .16920)
          /gene="C30B5.5"
CDS       complement(join(10349. .10585,10628. .10734,11184. .11582,
11628. .11781,12053. .12141,12885. .13153,13435. .13689,
14032. .14232,14680. .14807,14991. .15107,15299. .15380,
16335. .16490,16742. .16920))
          /gene="C30B5.5"
          /note="similar to adrenergic receptor"
          /codon_start=1
          /evidence=not_experimental
          /protein_id="AAK31469.1"
          /db_xref="GI:13592370"
          /translation="MDVIGNITDLSPTVSGIPDECGLPHDFLEVKFFLISVVGTLIG
LFGFLGNATTALILTRPSMRNPNNLFLTALAVFDSCLLITAFFIYAMEYII EYTAAFD
LYVAWLTYLRFAFALSHISQTGSVYITVAVTIERYLAEF SRLIFQVTVNPSCPDGSNW
QSYILLPSAMASNPIYQQVYSLWVTNFMVFFPFLTLLLFN AIIAYTIRQSLEKYDFH
NQKSVVAALSASVNLPRNIAGYCPSDCLCHVPLSSFTLRGLAPRRTVVYDCSRGSASS
VATSLTSPESFRISSRNELKEKSREATLVLVII VFIFLGCNFWG FVLTLLERIMGQET
LMVEHHIFYTFSREAINFLAIINSSINFVIYLLFGKDFR KELVVVYGCGIRGISLRLP
VQDKFVIWRHWKRTKSRISMNTTNRTRHKISLPQTLVEHANLERLEETRFLAHHEDGV
QTQVSPihalRNGSTPKIDTLQDLTSNGRPCKTSIIDDNGTVGKPRCNAFSIKIYLSL
FNAQKMKILVNILILIRITSGMLYDGKYFDMKDRIRAECIIKKAMSDYKYLGNNDKSV
NGEPCVPWIEVTESWLPTASDSQKMKKQASESFHHSKCRNIKLSVVNPMYSVSTPSAI
ETGIPSGIHGPWCFIDKVG TNGTESFKYTPVTCFN YCDETKVATDNEKMRLTENG YAI
LKQNYNPALLDPIEKLFT EYQFGNMKYTTFKKSREQPPQYLELRQKVFIALCLVFCV
IAWILSCYFLKKHSQKLHQKKQKALEGFYESNNLADIKLQADLRREREEA"
gene      complement(27490. .31388)
          /gene="C30B5.6"
CDS       complement(join(27490. .27593,27675. .27782,27835. .27927,
27971. .28064,28802. .29630,29761. .30035,30079. .30218,
30288. .30444,31125. .31247,31308. .31388))
          /gene="C30B5.6"
          /codon_start=1
          /evidence=not_experimental
          /product="Hypothetical protein C30B5.6"
          /protein_id="AAK31470.1"
          /db_xref="GI:13592371"
          /translation="MSVSTAFLLLLWSSIAGSLPLAPASPDFKADFSKLLKEAHQTDL
MPVVEEDELQSLRHNGFQVFYTFHEQLHHLRQKTVGRNSATITEEMARIRNSIEDISE
FSLDSNEVDEAINDITVPYGT KMKKSSKFAELSEIHNTDSSNKALPPFAFSEDQ QESV

```

```

Query Match      8.9%;   Score 39.4;   DB 3;   Length 41334;
Best Local Similarity  47.7%;   Pred. No. 0.49;
Matches 115;   Conservative 0;   Mismatches 126;   Indels 0;   Gaps 0;

```

```

Qy      126 ggcaaaccaaagaggacagcaatgctaggaaaatgacgatgacaaagacgagggcgatcgg 185
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      5156 GGAAGAATATAAAGTCCCGAAATACAAAGAAGATGCCGATGATGAAACGAAACGATTATG 5097

```

Qy 186 gcaacatacttggttagccgtaatgacgacgggccatgctatctagattccggtcttaatg 245  
 | | | | | | | | | | | | | | | | | | | | | |  
 Db 5096 GGAAGAAGGATGTGCTCCAAAACAGTAATGAGAGAAGCAGCACCTATGGAAGTTCAAGA 5037

Qy 246 agtacgtctgcagaaagactaataagtgcataagagcttggtgctctgcgtggcgagtt 305  
 | | | | | | | | | | | | | | | | | | | | | |  
 Db 5036 ACAGAGAATCAAGAAAGCCAAAGAAGTATTATTAGATATTGGAGATGTTGATGAAGAGTT 4977

Qy 306 gtcaaccatcatcatgaattcaagatactgcggagacatcatgatactgcggagacagac 365  
 | | | | | | | | | | | | | | | | | | | | | |  
 Db 4976 GCTGAAAAAATCAAAAAGATAAGAAGAAAGCAAAAAGGAGAAGAAGCGGGAAAAGAA 4917

Qy 366 g 366  
 |  
 Db 4916 G 4916

RESULT 7  
 AC080021  
 LOCUS AC080021 188254 bp DNA HTG 03-FEB-2001  
 DEFINITION Mus musculus clone RP23-422L7, WORKING DRAFT SEQUENCE, 12 unordered pieces.  
 ACCESSION AC080021  
 VERSION AC080021.2 GI:11138185  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 188254)  
 AUTHORS McCombie,W.R., Baker,J.P., Bahret,A., Bal,H., Baliya,V., Dedhia,N.N., de la Bastide,M., Huang,E.N., King,L., Kirchoff,K.A., Miller,B., Nascimento,L.U., O'Shaughnessy,A.L., Preston,R.R., Rodriguez,S., Santos,L., Shah,R.S., Spiegel,L.A., Toth,K., Vil,M.D. and Zutavern,T.  
 TITLE Mouse Genomic Sequence  
 JOURNAL Unpublished

REFERENCE 2 (bases 1 to 188254)  
 AUTHORS McCombie,W.R.  
 TITLE Direct Submission  
 JOURNAL Submitted (23-SEP-2000) Lita Annenberg Hazen Genome Sequencing Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA

COMMENT On Nov 11, 2000 this sequence version replaced gi:10280739.  
 ----- Genome Center  
 Center: Lita Annenberg Hazen Genome Center, Cold Spring Harbor Laboratory  
 Center code: CSHL  
 Web site: <http://www.cshl.org/genseq>  
 Contact: [mccombie@cshl.org](mailto:mccombie@cshl.org)  
 ----- Project Information  
 Center project name: RP23-422L7  
 Center clone name: RP23-422L7  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 12 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is



DEFINITION Homo sapiens chromosome 2 clone RP11-461M18, WORKING DRAFT  
SEQUENCE, 5 unordered pieces.

ACCESSION AC020602

VERSION AC020602.5 GI:14029092

KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ACTIVEFIN.

SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 170225)

AUTHORS Waterston,R.H.

TITLE The sequence of Homo sapiens clone

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 170225)

AUTHORS Waterston,R.H.

TITLE Direct Submission

JOURNAL Submitted (05-JAN-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA

COMMENT On May 12, 2001 this sequence version replaced gi:13992766.

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC

Web site:<http://genome.wustl.edu/gsc/index.shtml>

----- Project Information -----

Center project name: H\_NH0461M18

----- Summary Statistics -----

Sequencing vector: M13; 65%

Sequencing vector: plasmid; 31%

Chemistry: Dye-primer ET; 65% of reads

Chemistry: Dye-terminator Big Dye; 31% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 166879 bases at least Q40

Consensus quality: 167963 bases at least Q30

Consensus quality: 168640 bases at least Q20

Insert size: 166000; agarose-fp

Insert size: 169825; sum-of-contigs

Quality coverage: 7.42 in Q20 bases; agarose-fp

Quality coverage: 7.36 in Q20 bases; sum-of-contigs

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 5 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\*       1       1167: contig of 1167 bp in length  
\*     1168       1267: gap of unknown length  
\*     1268       2828: contig of 1561 bp in length  
\*     2829       2928: gap of unknown length  
\*     2929       36359: contig of 33431 bp in length  
\*     36360       36459: gap of unknown length  
\*     36460       86624: contig of 50165 bp in length  
\*     86625       86724: gap of unknown length





REFERENCE 2 (bases 1 to 49307)  
AUTHORS Waterston, R.H.  
TITLE Direct Submission  
JOURNAL Submitted (01-MAR-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA

REFERENCE 3 (bases 1 to 49307)  
AUTHORS Waterston, R.  
TITLE Direct Submission  
JOURNAL Submitted (24-MAR-2000) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

REFERENCE 4 (bases 1 to 49307)  
AUTHORS Waterston, R.  
TITLE Direct Submission  
JOURNAL Submitted (06-APR-2001) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

COMMENT Submitted by:  
Genome Sequencing Center  
Department of Genetics, Washington University,  
St. Louis, MO 63110, USA, and  
Sanger Centre, Hinxton Hall  
Cambridge CB10 1RQ, England  
e-mail: rw@nematode.wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.  
It may be shorter because we only sequence overlapping sections  
once, or longer because we provide a small overlap between  
neighboring submissions.

#### NOTES:

Coding sequences below are predicted from computer analysis, using  
the program Genefinder (P. Green and L. Hillier, ms in preparation).

FEATURES Location/Qualifiers

source	1. .49307 /organism="Caenorhabditis elegans" /strain="Bristol N2" /db_xref="taxon:6239" /chromosome="I" /clone="CELY34D9A"
gene	complement(1791. .11784) /gene="Y34D9A.1"
CDS	complement(join(1791. .1910,5779. .6091,8043. .8493, 11369. .11479,11538. .11784)) /gene="Y34D9A.1" /note="coded for by C. elegans cDNA yk56g7.3; coded for by C. elegans cDNA yk69f5.5; coded for by C. elegans cDNA yk56g7.5; coded for by C. elegans cDNA yk632g7.5" /codon_start=1 /product="Hypothetical protein Y34D9A.1" /protein_id="AAK29884.1" /db_xref="GI:13559675" /translation="MSAAQYARLVPKKYRSKTLPKIDRPWRPRVIAWAGPAAFYPNRF YEVDKWYKARIDKPEKLPEMHIEPAEHMKSLSKVLQKSEIEQINIGFKRREVAGKAE KSAEDRIELERKSRHMQKIDIDHLDVENLSIYRHFQVFDHLFGDNIFFENVQNLQVN"

```

FENDIVVHSGNVITANSTLKRPEITIESVGNNGGFNTLLMINLDGNALDLGKNGEIVQ
WMISNIPDGEAISAGSEIIDYLQPLPFYGTGYHRVAFVLFREHEKPVDFQIQGNSLDTR
IHEISKFYKKHEATITPSAIRFFQTSYDNSVKMALHGLGMTSPLYEYEHRLPALKPAQR
EFPEKPQPFDFLYLDMYRDPKEVEQEEMLEKRLAEVKLDYVKEPKWVDTDYVENKKKLPA
WLHAKKLERDGVGHAKYHNDL"
gene      complement(12705. .15311)
          /gene="Y34D9A.4"
CDS       complement(join(12705. .12959,14804. .15088,15141. .15311))
          /gene="Y34D9A.4"
          /note="coded for by C. elegans cDNA yk747h4.5; coded for
          by C. elegans cDNA yk110f3.5; coded for by C. elegans cDNA
          yk82a1.5; coded for by C. elegans cDNA yk267c3.5; coded
          for by C. elegans cDNA yk461f9.5; coded for by C. elegans
          cDNA yk425b7.5; coded for by C. elegans cDNA yk292d8.5;
          coded for by C. elegans cDNA yk533f3.5; coded for by C.
          elegans cDNA yk470a4.5"
          /codon_start=1
          /product="Hypothetical protein Y34D9A.4"
          /protein_id="AAK29885.1"
          /db_xref="GI:13559676"
          /translation="MSRRHSVVDAYSTVEGKIRDTMRELSALWDEVDMSEAMRLKRVD
          NAFTHITLLCDDMLSGEKEMIHNLKVSIREDMQNVKKMRLELEMEDFQRP AEIKDGS I
          ALMRHLQSEVKSLETEFQTRHEDQRVLIEKICNLKKRLESDFEF EY EIHAKLNFDTDL
          PPCMSPATSEMISFITPTRRQAVGPKTSSPKDTQPSTSSSSR TTPMSKRAMPSSIA
          SSTPSSAKKVLTRRNQFL"
gene      14384. .22639
          /gene="Y34D9A.3"
CDS       join(14384. .14680,15517. .15594,16343. .16373,16423.
.16520,
          16567. .16755,16808. .16939,16997. .17134,21214. .21757,
          21990. .22522,22574. .22639)
          /gene="Y34D9A.3"
          /note="coded for by C. elegans cDNA yk533f3.5; coded for
          by C. elegans cDNA yk747h4.5; coded for by C. elegans cDNA
          yk35d4.5; coded for by C. elegans cDNA yk154c9.5"
          /codon_start=1
          /product="Hypothetical protein Y34D9A.3"
          /protein_id="AAK29886.1"
          /db_xref="GI:13559677"
          /translation="MSICFRYETFLPKLLKLLMLLGPSEL DIFQ PMLVKRLQLAVVRL
          QRVQKPLREHKFRGIDVLLQQILQIIIPADIVGHILRHTLPLRDVGL ETLDLVVAFFV
          ARPVPEENACAFNFSLILNIFEQQIMRRRRRCIPVINDITEGFAAE ELEYAVFDES LFT
          SIWKSPERLSNETDFHLLLTALSREIAAENYQEPLLILKFMKKTAIGQDEVINRIFTS
          LLKQILRDETRKKLDPHVLD FLDAGDKLFLDSETSKKAYAILKLYFLCRNRKSKLEC
          FSCGCVQKYFIELYKDGTRLHEALNCLI KSKLSDSPTI IERKPEKR I I I I FFDIMKLL
          GAEKPKIHSYFKENRTKTKLLMNSELELAVEYLRDFSDYELNWFLOQRADEDFQFLLE
          TAKNLQAPVNSDGFWRNSLEKFDENSII FIENQLKIMATRIDRKG LALRIQQISQKMM
          QKSGKIEEKTAIFIKFPQILLEIADFSILPPNSSNLLFSTLILGFSAFYDGNEAQFND
          LFRQFLSKTIPNSTCTSPMEIIEEICEKIRIKSNFRHL PDISFLKPNPDTIITQISEI
          FDTKIAKNDERIVEICSI LLEKLAPQLQAFGDRGKTPFCQIVNIIVDKYVKNSVEPVN
          FQLFREVCERAVDGFAMMTEQWNFLFLSAKILTILNLTRREAGLFAVLMRKINKNEIL
          LTKLRGKKEFVQLEGSLEL"
gene      23360. .30062
          /gene="Y34D9A.7"
CDS       join(23360. .23566,25232. .25801,27316. .27378,28218.
.28454,
          29198. .29291,29380. .29552,29610. .29687,29832. .30062)

```

```

/gene="Y34D9A.7"
/codon_start=1
/evidence=not_experimental
/product="Hypothetical protein Y34D9A.7"
/protein_id="AAK29880.1"
/db_xref="GI:13559671"
/translation="MSEDGEISLSPNSAALMDGLDDFIIDKNCSGII PNEDLERETPS
ADDDAEVKTSSNWAKNMRKRMFGEQCDEPEESEKKRRRTNCCFNCRGGDHSIAQCPEP
KNFAEIRKKNKLEFMNDKQQQHQP TGRISQVTEQQQEAKFKPGRLSQNLRKALS LGPDD
IPEWVYRMRLGFYRGYP PGYLRKSLKREFATLKIYSEDHNQEEVDNDDDEEARPAPT
IQSEKVHFYMGFNKTYGALRDRERGRFEVPPFDIFCEMLQTFLPNSYLNFRQLDLFDG
ARLHEVARDHEGTEKIRIREERHQ RSELRKIREQQKEEEEQQQORDSAEKVEKEVTVDQ
TISEVAAAAAAVEEKP GTPEPVGRMKKLNFLNKKLIFFSRMLYFSKIKSKFYSRKKVK
KFFFSKNQYSSKKLK FSSISLPKIYISRISISNFFFKVFCPKIDFFFPKFFQIFFPKT
YFSKKKKRIRKSIAGILLNTGKFPGESISEFIGTPIFSRRDVTGTWIEDTVPSLEAFS
VGIVPFEAKEEEKPRGIFKKIMSTLKGITGRNSGDDEMKKKE"
gene 31479. .32514
/gene="Y34D9A.6"
CDS join(31479. .31562,32281. .32514)
/gene="Y34D9A.6"
/codon_start=1
/evidence=not_experimental
/product="Hypothetical protein Y34D9A.6"
/protein_id="AAK29881.1"
/db_xref="GI:13559672"
/translation="MSKAFVDGLLQSSKVVVFSKSYCPYCHKARAALSVNVKPDALQ
WIEIDERKDCNEIQDYLGS LTGARSVPRVFINGKFFGGGDDTAAGAKNGKLAALLKET
GAL"
gene complement(32800. .34038)
/gene="Y34D9A.8"
CDS complement(join(32800. .33005,33792. .33901,33968. .34038))
/gene="Y34D9A.8"
/codon_start=1
/evidence=not_experimental
/product="Hypothetical protein Y34D9A.8"
/protein_id="AAK29879.1"
/db_xref="GI:13559670"
/translation="MADSLLSNIIQQEITDFPELFD MGGGAPMAREGVAQPRQTNVQA
TVA AVKETTTITAESSGTVTIQYSHIFFAFVAFFVLSVAVVAVIRRRSRQKSGFRNR
GGGHGGPSILQQDSEDDILISSMYS"
gene complement(36238. .44285)
/gene="Y34D9A.9"
CDS complement(join(36238. .36345,37937. .38081,39570. .39817,
40470. .40532,42326. .42546,43973. .44285))
/gene="Y34D9A.9"
/codon_start=1
/evidence=not_experimental
/product="Hypothetical protein Y34D9A.9"
/protein_id="AAK29878.1"
/db_xref="GI:13559669"
/translation="MSEYEPIGIDYTH THTQTLRSSLQMTPCGSQRQGEKRREKKWDR
ECFCVCVSCCVSRCWDHTTC FCCSSAALLSHTLLRLSCSSLLSAAPHTLELGGGSLS
SSGNRKPWQGILLFGPPGTGKSYIAKAVATEAGESTFFS ISSSDLMSKWLGESEKLVK
NLFALAREHKPSIIFIDEFLPNSYSNFRQSALFDGARLHQFNTYFSNNRRFEKRIYIP
LPDIHARKEMFRIDVGKNYNTLT DQDFKVLAERCEGYSGYDISILVKDALMQPVRVQ
SATHFKHVSGPSPKDPNVIA HDLLTPCSPGDPHAIAMNWLDPGDKLANPPLSMQDIS
RSLASVKPTVNNTDLDRLEAFKNDFGQDQGE"

```

gene 44323. .46794  
 /gene="Y34D9A.2"  
 CDS join(44323. .44706,45476. .45882,46530. .46794)  
 /gene="Y34D9A.2"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /product="Hypothetical protein Y34D9A.2"  
 /protein\_id="AAK29882.1"  
 /db\_xref="GI:13559673"  
 /translation="MYGFFNSSEIDEEYYNSTHTSAPSPILALIFTIICIIGVIGNAS  
 LLVYIFAKKLYQNFISRFIGHLCTNLIALLVLPVVIHNVFTGVNLLQDSNMLCRI  
 QVTETFSAVTWSQKVPFRFDLRKMREVSITVTVWTVIAMMNLICIAGVHLLTFARIHYE  
 QLFGLTPTKLCILSWIISWLLSLPSLTNGHVAIYGPAVRTCVFVSHSDSGLKFLTYTMI  
 FGVFIPALFSSIAFYFRILQTLFHSPIVFQSLGLYKSRFLVYFFLLGPLYALPFYILTA

Query Match 8.1%; Score 36; DB 3; Length 49307;  
 Best Local Similarity 47.0%; Pred. No. 5.6;  
 Matches 111; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

Qy 96 ccttggtgcccgaagcaagccacaaaaagtgggcaaaccgaagaggacagcaatgctagga 155  
 ||||| | | || | | || | | || | | || | | ||  
 Db 16858 CCTTGATTCTGAAACTTCAAAAAAGCATATGCAATTCTAAAATTGTATTTTCTCTGCAG 16917  
 Qy 156 aaatgacgatgacaaagacgagggcatcgggcaacatacttgtagccgtaatgacgacg 215  
 |||| | || | || | | | | | | | | | | || ||  
 Db 16918 AAATCGAAAATCCAAGCTAGAGGTTGGTGATTTTCCGCCCGGAAAACCTGAAAACCCGCCG 16977  
 Qy 216 ggccatgctatctagattccggtcttaatgagtacgtctgcagaaagactaataagtgct 275  
 | | | | | | | | | | | | | | | | | | | |  
 Db 16978 AAAATTCATTTTTTCCAGTGTTTTTCGTGCGGTTGTGTGCAAAATACTTTATCGAGCT 17037  
 Qy 276 ataagagcttggtgctctgcgtggcgagttgtcaaccatcatcatgaattcaagat 331  
 || || |||| | || | |||| | || | || || || || ||  
 Db 17038 CTACAAGGATGGTACGAGGCTTCACGAGGCGTTAAACTGCCTGATAAAATCAAAAT 17093

RESULT 10  
 AC006735/c  
 LOCUS AC006735 166214 bp DNA HTG 25-FEB-1999  
 DEFINITION Caenorhabditis elegans clone Y34D9, \*\*\* SEQUENCING IN PROGRESS \*\*\*,  
 4 unordered pieces.  
 ACCESSION AC006735  
 VERSION AC006735.3 GI:4309801  
 KEYWORDS HTG; HTGS\_PHASE1.  
 SOURCE Caenorhabditis elegans.  
 ORGANISM Caenorhabditis elegans  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;  
 Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  
 REFERENCE 1 (bases 1 to 166214)  
 AUTHORS Waterston,R.H.  
 TITLE The sequence of Caenorhabditis elegans clone  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 166214)  
 AUTHORS Waterston,R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (23-FEB-1999) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

COMMENT On Mar 1, 1999 this sequence version replaced gi:4263429.  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 4 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1 2055: contig of 2055 bp in length  
 \* 2056 2069: gap of unknown length  
 \* 2070 20299: contig of 18230 bp in length  
 \* 20300 20313: gap of unknown length  
 \* 20314 70328: contig of 50015 bp in length  
 \* 70329 70342: gap of unknown length  
 \* 70343 166214: contig of 95872 bp in length.

FEATURES Location/Qualifiers  
 source 1. .166214  
 /organism="Caenorhabditis elegans"  
 /db\_xref="taxon:6239"  
 /clone="Y34D9"

BASE COUNT 52633 a 31045 c 31031 g 51463 t 42 others  
 ORIGIN

Query Match 8.1%; Score 36; DB 2; Length 166214;  
 Best Local Similarity 47.0%; Pred. No. 6.4;  
 Matches 111; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

Qy 96 ccttggtgcccaagcaagccacaaaaagtgggcaaaccagaggacagcaatgctagga 155  
 ||||| | | || | | || | | || | | || | | ||  
 Db 81072 CCTTGATTCTGAAACTTCAAAAAAGCATATGCAATTCTAAAATTGTATTTTCTCTGCAG 81013

Qy 156 aaatgacgatgacaaagacgagggcatcgggcaacatacttggttagccgtaatgacgacg 215  
 |||| | | || | || | | | | | | | | | | | |  
 Db 81012 AAATCGAAAATCCAAGCTAGAGGTTGGTGATTTTCCGCCCGGAAACTGAAAACCCGCCG 80953

Qy 216 ggccatgctatctagattccggtcttaatgagtacgtctgcagaaagactaataagtgct 275  
 | | | | | | | | | | | | | | | | | | | |  
 Db 80952 AAAATTCCATTTTTTCCAGTGTTTTTCGTGCGGTTGTGTGCAAAAATACTTTATCGAGCT 80893

Qy 276 ataagagcttggtgctctgctggtggcgagttgtcaaccatcatcatgaattcaagat 331  
 || || |||| | || | |||| || || | | | | | | | |  
 Db 80892 CTACAAGGATGGTACGAGGCTTCACGAGGCGTTAAACTGCCTGATAAAATCAAAAT 80837

RESULT 11  
 AC011081  
 LOCUS AC011081 183980 bp DNA HTG 10-SEP-2000  
 DEFINITION Homo sapiens clone RP11-45019, WORKING DRAFT SEQUENCE, 20 unordered  
 pieces.  
 ACCESSION AC011081  
 VERSION AC011081.3 GI:10047669  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE human.  
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 183980)  
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
TITLE Homo sapiens, clone RP11-45019  
JOURNAL Unpublished

REFERENCE 2 (bases 1 to 183980)  
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,  
Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,  
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,  
Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,  
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,  
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,  
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,  
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,  
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,  
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

TITLE Direct Submission  
JOURNAL Submitted (01-OCT-1999) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT On Sep 10, 2000 this sequence version replaced gi:7637227.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L1215  
Center clone name: 45\_O\_19  
----- Summary Statistics  
Sequencing vector: M13; M77815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 165551 bases at least Q40  
Consensus quality: 173368 bases at least Q30  
Consensus quality: 177359 bases at least Q20  
Insert size: 198000; agarose-fp  
Insert size: 182080; sum-of-contigs  
Quality coverage: 3.5 in Q20 bases; agarose-fp  
Quality coverage: 3.8 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 20 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 5302: contig of 5302 bp in length

```

*      5303 5402: gap of      100 bp
*      5403      6656: contig of 1254 bp in length
*      6657 6756: gap of      100 bp
*      6757      8184: contig of 1428 bp in length
*      8185 8284: gap of      100 bp
*      8285      10720: contig of 2436 bp in length
*     10721 10820: gap of      100 bp
*     10821      13684: contig of 2864 bp in length
*     13685 13784: gap of      100 bp
*     13785      15838: contig of 2054 bp in length
*     15839 15938: gap of      100 bp
*     15939      17986: contig of 2048 bp in length
*     17987 18086: gap of      100 bp
*     18087      21602: contig of 3516 bp in length
*     21603 21702: gap of      100 bp
*     21703      24984: contig of 3282 bp in length
*     24985 25084: gap of      100 bp
*     25085      29305: contig of 4221 bp in length
*     29306 29405: gap of      100 bp
*     29406      57392: contig of 27987 bp in length
*     57393 57492: gap of      100 bp
*     57493      64118: contig of 6626 bp in length
*     64119 64218: gap of      100 bp
*     64219      71303: contig of 7085 bp in length
*     71304 71403: gap of      100 bp
*     71404      79950: contig of 8547 bp in length
*     79951 80050: gap of      100 bp
*     80051      91968: contig of 11918 bp in length
*     91969 92068: gap of      100 bp
*     92069      107529: contig of 15461 bp in length
*    107530 107629: gap of      100 bp
*    107630      128021: contig of 20392 bp in length
*    128022 128121: gap of      100 bp
*    128122      148757: contig of 20636 bp in length
*    148758 148857: gap of      100 bp
*    148858      178744: contig of 29887 bp in length
*    178745 178844: gap of      100 bp
*    178845      183980: contig of 5136 bp in length.

```

```

FEATURES              Location/Qualifiers
  source              1. .183980
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="RP11-45O19"
                      /clone_lib="RPCI-11 Human Male BAC"
  misc_feature        1. .5302
                      /note="assembly_fragment"
                      clone_end:SP6
                      vector_side:left"
  misc_feature        5403. .6656
                      /note="assembly_fragment"
  misc_feature        6757. .8184
                      /note="assembly_fragment"
  misc_feature        8285. .10720
                      /note="assembly_fragment"
  misc_feature        10821. .13684
                      /note="assembly_fragment"
  misc_feature        13785. .15838

```

```

misc_feature      /note="assembly_fragment"
15939. .17986
misc_feature      /note="assembly_fragment"
18087. .21602
misc_feature      /note="assembly_fragment"
21703. .24984
misc_feature      /note="assembly_fragment"
25085. .29305
misc_feature      /note="assembly_fragment"
29406. .57392
misc_feature      /note="assembly_fragment"
57493. .64118
misc_feature      /note="assembly_fragment"
64219. .71303
misc_feature      /note="assembly_fragment"
71404. .79950
misc_feature      /note="assembly_fragment"
80051. .91968
misc_feature      /note="assembly_fragment"
92069. .107529
misc_feature      /note="assembly_fragment"
107630. .128021
misc_feature      /note="assembly_fragment"
128122. .148757
misc_feature      /note="assembly_fragment"
148858. .178744
misc_feature      /note="assembly_fragment"
178845. .183980
misc_feature      /note="assembly_fragment
clone_end:T7
vector_side:right"

```

BASE COUNT 57357 a 33536 c 32891 g 58294 t 1902 others  
ORIGIN

Query Match 8.1%; Score 36; DB 2; Length 183980;  
Best Local Similarity 56.9%; Pred. No. 6.5;  
Matches 66; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

```

Qy      46 aacaacctccaaagaataatccgggtgccttccaagaatcctccaaccacccttggtgcc 105
      |||| | ||  ||  | || |||| ||  ||  | ||| | ||| || || ||
Db 25914 AACATCTTCAGTAGGCCAGCCCGGTGGCTCACACCTGTAATCCCAGCACTCTGGGAGGC 25973

Qy      106 caagcaagccacaaaaagtgggcaaaccaaagaggacagcaatgctaggaaaatga 161
      | || ||| || | | | || || | | ||| ||| | ||| ||| |||
Db 25974 CGAGGAAGGCAGATCACGAGGTCATCCGATCGAGACCATCCTGGCTAACACAGTGA 26029

```

RESULT 12  
AL353592/c  
LOCUS AL353592 63325 bp DNA PRI 30-MAY-2001  
DEFINITION Human DNA sequence from clone RP11-569012 on chromosome 13,  
complete sequence.  
ACCESSION AL353592  
VERSION AL353592.9 GI:14272259  
KEYWORDS HTG.  
SOURCE human.



ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 63325)

AUTHORS Sycamore, N.

TITLE Direct Submission

JOURNAL Submitted (30-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone  
 requests: clonerequest@sanger.ac.uk

COMMENT On May 31, 2001 this sequence version replaced gi:13751329.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >= 30);  
 an attempt was made to resolve all sequencing problems, such as  
 compressions and repeats; all regions were covered by at least one  
 plasmid subclone or more than one M13 subclone; and the assembly was  
 confirmed by restriction digest. The following abbreviations are used to  
 associate primary accession numbers given in the feature table with their  
 source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP;  
 Information on the WORMPEP database can be found at  
[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence was  
 generated from part of bacterial clone contigs of human chromosome 13,  
 constructed by the Sanger Centre Chromosome 13 Mapping Group. Further  
 information can be found at <http://www.sanger.ac.uk/HGP/Chr13>  
 RP11-569012 is from the library RPCI-11.2 constructed by the group of  
 Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>  
 VECTOR: pBACe3.6  
 IMPORTANT: This sequence is not the entire insert of clone RP11-569012.  
 It may be shorter because we sequence overlapping sections only once,  
 except for a 100 base overlap. The true left end of clone RP11-274M5 is  
 at 63226 in this sequence. The true right end of clone RP11-431P10 is  
 at 100 in this sequence.

FEATURES

	Location/Qualifiers
source	1..63325 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="13" /clone="RP11-569012" /clone_lib="RPCI-11.2"
repeat_region	444..814 /note="MER7A repeat: matches 1..345 of consensus"
repeat_region	1032..1121 /note="MIR repeat: matches 80..165 of consensus"
repeat_region	2068..2169 /note="MIR repeat: matches 73..177 of consensus"
repeat_region	6485..6506 /note="11 copies 2 mer aa 100% conserved"
repeat_region	8486...8616

	/note="MER94 repeat: matches 5. .134 of consensus"
repeat_region	8809. .8956
	/note="MIR repeat: matches 6. .168 of consensus"
repeat_region	8958. .9043
	/note="THE1C repeat: matches 279. .371 of consensus"
repeat_region	9044. .9794
	/note="THE1B-INTERNAL repeat: matches 1. .754 of consensus"
repeat_region	9795. .10142
	/note="THE1B repeat: matches 1. .364 of consensus"
repeat_region	10148. .10217
	/note="MIR repeat: matches 162. .230 of consensus"
repeat_region	13570. .14052
	/note="HAL1 repeat: matches 170. .660 of consensus"
repeat_region	14063. .14360
	/note="HAL1 repeat: matches 627. .915 of consensus"
repeat_region	14386. .14485
	/note="MER94 repeat: matches 33. .134 of consensus"
repeat_region	15882. .15929
	/note="24 copies 2 mer tt 75% conserved"
repeat_region	18054. .18355
	/note="AluSq repeat: matches 11. .310 of consensus"
repeat_region	22941. .23051
	/note="AluJb repeat: matches 5. .120 of consensus"
repeat_region	24374. .24492
	/note="MIR repeat: matches 137. .259 of consensus"
repeat_region	24746. .24799
	/note="27 copies 2 mer aa 75% conserved"
repeat_region	25942. .26323
	/note="THE1C repeat: matches 3. .371 of consensus"
repeat_region	28358. .30231
	/note="L1PA16 repeat: matches 4376. .6157 of consensus"
repeat_region	30232. .30589
	/note="L1PA4 repeat: matches 5789. .6146 of consensus"
repeat_region	30590. .32110
	/note="L1PA16 repeat: matches 2931. .4376 of consensus"
repeat_region	32111. .32533
	/note="MLT2B repeat: matches 1. .444 of consensus"
repeat_region	32534. .34068
	/note="L1PA16 repeat: matches 1075. .2931 of consensus"
repeat_region	34076. .35640
	/note="L1PA4 repeat: matches 4576. .6140 of consensus"
repeat_region	35646. .36278
	/note="L1PA15-16 repeat: matches 476. .1121 of consensus"
repeat_region	36273. .36337
	/note="L1PA15-16 repeat: matches 495. .559 of consensus"
repeat_region	36329. .37521
	/note="L1PA15-16 repeat: matches -694. .492 of consensus"
repeat_region	38106. .38757
	/note="L1PA13 repeat: matches 5533. .6156 of consensus"
repeat_region	38772. .38885
	/note="FLAM_C repeat: matches 1. .123 of consensus"
repeat_region	41019. .41398
	/note="L1ME3 repeat: matches 5768. .6148 of consensus"
repeat_region	41706. .41745
	/note="20 copies 2 mer tg 80% conserved"
repeat_region	43615. .44111
	/note="MER1A repeat: matches 1. .527 of consensus"

repeat\_region 44388. .44538  
 /note="MIR repeat: matches 92. .249 of consensus"  
 repeat\_region 46500. .46741  
 /note="MIR repeat: matches 3. .258 of consensus"  
 repeat\_region 48865. .49034  
 /note="MIR repeat: matches 79. .247 of consensus"  
 repeat\_region 49037. .49406  
 /note="MLT1A2 repeat: matches 1. .370 of consensus"  
 repeat\_region 49889. .50199  
 /note="AluY repeat: matches 1. .309 of consensus"  
 repeat\_region 50711. .50738  
 /note="14 copies 2 mer ct 92% conserved"  
 repeat\_region 53571. .53666  
 /note="8 copies 12 mer 66% conserved"  
 repeat\_region 53672. .53975  
 /note="AluSx repeat: matches 1. .305 of consensus"  
 repeat\_region 54085. .54655  
 /note="MER67C repeat: matches 122. .709 of consensus"  
 repeat\_region 54985. .55043  
 /note="MER5A repeat: matches 103. .162 of consensus"  
 repeat\_region 54995. .55073  
 /note="MER5A repeat: matches 9. .88 of consensus"  
 repeat\_region 55210. .55514  
 /note="L1PA6 repeat: matches 5837. .6142 of consensus"  
 repeat\_region 56363. .56729  
 /note="MLT1A1 repeat: matches 1. .365 of consensus"  
 repeat\_region 58149. .58262  
 /note="MIR repeat: matches 79. .196 of consensus"  
 repeat\_region 58442. .58489  
 /note="4 copies 12 mer 89% conserved"  
 repeat\_region 60568. .60629  
 /note="L1MA9 repeat: matches 6110. .6158 of consensus"  
 repeat\_region 60630. .61116  
 /note="MER1A repeat: matches 1. .527 of consensus"  
 repeat\_region 61117. .61160  
 /note="L1MA9 repeat: matches 6064. .6110 of consensus"  
 repeat\_region 61152. .61349  
 /note="L1MA10 repeat: matches 6140. .6320 of consensus"  
 repeat\_region 61527. .61634  
 /note="MIR repeat: matches 51. .160 of consensus"  
 repeat\_region 61658. .61711  
 /note="MER92B repeat: matches 4. .58 of consensus"

BASE COUNT 19347 a 11930 c 11644 g 20404 t  
 ORIGIN

Query Match 8.0%; Score 35.6; DB 9; Length 63325;  
 Best Local Similarity 58.5%; Pred. No. 7.6;  
 Matches 62; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

Qy 220 atgctatctagattccggtcttaatgagtagctgtgcagaaagactaataagtgtctataa 279  
 || || ||||| | | ||||| | || ||||| ||||| | |||  
 Db 11138 ATATAATGTAGATTAAGATATTAATGAATTTGTTTTTCAGAATCTGTAATAGGGACCCTAA 11079  
  
 Qy 280 gagcttggtgctctgctgtggcgagttgtcaaccatcatcatgaatt 325  
 |||| ||||| | | || || ||||| ||||  
 Db 11078 GAGCAGGGTGCTATCATTACAACTCTTCCCTTTTATTACTAATT 11033

RESULT 13

ZMA297901

LOCUS ZMA297901 563 bp mRNA PLN 11-JAN-2001

DEFINITION Zea mays mRNA for basal layer antifungal peptide (bap-1a gene).

ACCESSION AJ297901

VERSION AJ297901.1 GI:12214246

KEYWORDS bap-1a gene; basal layer antifungal peptide.

SOURCE Zea mays.

ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 563)

AUTHORS Serna Sanz, A. and Thompson, R.D.

TITLE Maize endosperm secretes a novel antifungal protein into adjacent  
maternal tissue

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 563)

AUTHORS Serna, A.

TITLE Direct Submission

JOURNAL Submitted (31-OCT-2000) Serna A., Plant Physiology, Max Planck  
Institut, Carl von Linne Weg 10, Cologne, 50829, GERMANY

FEATURES Location/Qualifiers

source 1..563

/organism="Zea mays"

/variety="A69Y"

/db\_xref="taxon:4577"

/tissue\_type="endosperm"

/dev\_stage="7 days after pollination"

sig\_peptide 40..123

/gene="bap-1a"

CDS 40..321

/gene="bap-1a"

/function="putative antifungal peptide"

/codon\_start=1

/product="basal layer antifungal peptide"

/protein\_id="CAC21605.1"

/db\_xref="GI:12214247"

/translation="MAKFFNYTIIQGLLMSVLLASCAIHAHIISGETEEVSNTGSP  
TVMVTMGANRKIIEDNKNLLCYLRALEYCCARTRQCYDDIKKCLEHCRG"

gene 40..321

/gene="bap-1a"

mat\_peptide 124..318

/gene="bap-1a"

/product="bap-1a protein"

BASE COUNT 201 a 95 c 109 g 158 t

ORIGIN

Query Match 8.0%; Score 35.4; DB 8; Length 563;

Best Local Similarity 49.0%; Pred. No. 5.1;

Matches 127; Conservative 0; Mismatches 126; Indels 6; Gaps 1;

Qy 50 acctccaaagaataatccgggtgccttccaagaatcctccaaccacccttggtgcccaag 109

||| || || | | | ||| | | | | | | | |

Db 61 ACCATCATCCAAGGACTCTTGATGCTTTCCATGGTACTTCTGGCATCGTGGCTATTTCAT 120  
 Qy 110 caagccacaaaaagtgggcaaaccaaagaggacagcaatgctaggaaaatgacgatgaca 169  
 | ||| ||| ||||| ||| ||||| ||||| | ||| ||| |||  
 Db 121 GCACACATAATAAGTGGGGAAACTGAAGAGGTTAGCAACACAGGGAGCCCCGACAGTGATG 180  
 Qy 170 aagacgagggcatcgggcaacatacttgtagccgtaatgacgacggggccatgctatcta 229  
 ||||| || | | | | ||| | | ||||| |||  
 Db 181 GTCACGATGGGGGCAAACCGAAAGATAATTGAAGATAATAAAAATTTATTGTGCTATCTA 240  
 Qy 230 gattccggtcttaaatgagtacgtctgcagaaagactaataagtgtataagagcttggtg 289  
 | | | | | ||||| || ||| || | ||||| | | | |  
 Db 241 AGGGC-----TCTAGAGTACTGTTGTGCAAGGACCAGACAATGCTATGATGACATAAAG 294  
 Qy 290 ctctgcgtggcgagttgtc 308  
 ||| ||| | ||| |  
 Db 295 AAATGCTTGGAGCATTGCC 313

# RESULT 14

AL355521

LOCUS AL355521 78874 bp DNA HTG 13-JUN-2001  
 DEFINITION Homo sapiens chromosome X clone RP11-723E19, \*\*\* SEQUENCING IN  
 PROGRESS \*\*\*, 22 unordered pieces.

ACCESSION AL355521

VERSION AL355521.4 GI:9863727

KEYWORDS HTG; HTGS\_PHASE1; HTGS\_CANCELLED.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 78874)

AUTHORS Mclay,K.

TITLE Direct Submission

JOURNAL Submitted (12-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone  
 requests: clonerequest@sanger.ac.uk

COMMENT On Aug 21, 2000 this sequence version replaced gi:9231037.

----- Genome Center

Center: Sanger Centre

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: humquery@sanger.ac.uk

----- Project Information

Center project name: bA723E19

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Sequencing vector: plasmid; L08752; 100% of reads

Chemistry: Dye-terminator ET-amersham; 100% of reads Consensus  
 quality: 62626 bases at least Q40

Consensus quality: 68358 bases at least Q30

Consensus quality: 72218 bases at least Q20

Insert size: 76774; sum-of-contigs

Insert size: 183906; agarose-fp

Quality coverage: 2.07x in Q20 bases; sum-of-contigs Quality  
 coverage: 1.51x in Q20 bases; agarose-fp

-----

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 22 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

```
*      1      2671: contig of 2671 bp in length
*      2672 2771: gap of      100 bp
*      2772      8329: contig of 5558 bp in length
*      8330 8429: gap of      100 bp
*      8430      11063: contig of 2634 bp in length
*     11064 11163: gap of      100 bp
*     11164      13181: contig of 2018 bp in length
*     13182 13281: gap of      100 bp
*     13282      16335: contig of 3054 bp in length
*     16336 16435: gap of      100 bp
*     16436      21088: contig of 4653 bp in length
*     21089 21188: gap of      100 bp
*     21189      23402: contig of 2214 bp in length
*     23403 23502: gap of      100 bp
*     23503      25575: contig of 2073 bp in length
*     25576 25675: gap of      100 bp
*     25676      32961: contig of 7286 bp in length
*     32962 33061: gap of      100 bp
*     33062      37451: contig of 4390 bp in length
*     37452 37551: gap of      100 bp
*     37552      40168: contig of 2617 bp in length
*     40169 40268: gap of      100 bp
*     40269      44181: contig of 3913 bp in length
*     44182 44281: gap of      100 bp
*     44282      49401: contig of 5120 bp in length
*     49402 49501: gap of      100 bp
*     49502      55754: contig of 6253 bp in length
*     55755 55854: gap of      100 bp
*     55855      58174: contig of 2320 bp in length
*     58175 58274: gap of      100 bp
*     58275      60640: contig of 2366 bp in length
*     60641 60740: gap of      100 bp
*     60741      64375: contig of 3635 bp in length
*     64376 64475: gap of      100 bp
*     64476      67814: contig of 3339 bp in length
*     67815 67914: gap of      100 bp
*     67915      70067: contig of 2153 bp in length
*     70068 70167: gap of      100 bp
*     70168      72968: contig of 2801 bp in length
*     72969 73068: gap of      100 bp
*     73069      75340: contig of 2272 bp in length
*     75341 75440: gap of      100 bp
*     75441      78874: contig of 3434 bp in length.
```

#### FEATURES

source

Location/Qualifiers

```
1. .78874
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/clone="RP11-723E19"
```

```

/clone_lib="RPCI-11.3"
misc_feature 1. .2671
/note="assembly_fragment:00756
fragment_chain:1"
misc_feature 2772. .8329
/note="assembly_fragment:00239
fragment_chain:1"
misc_feature 8430. .11063
/note="assembly_fragment:00798
fragment_chain:2"
misc_feature 11164. .13181
/note="assembly_fragment:00291
fragment_chain:2"
misc_feature 13282. .16335
/note="assembly_fragment:00029"
misc_feature 16436. .21088
/note="assembly_fragment:00042"
misc_feature 21189. .23402
/note="assembly_fragment:00139"
misc_feature 23503. .25575
/note="assembly_fragment:00155"
misc_feature 25676. .32961
/note="assembly_fragment:00215"
misc_feature 33062. .37451
/note="assembly_fragment:00227"
misc_feature 37552. .40168
/note="assembly_fragment:00290"
misc_feature 40269. .44181
/note="assembly_fragment:00333"
misc_feature 44282. .49401
/note="assembly_fragment:00387"
misc_feature 49502. .55754
/note="assembly_fragment:00398"
misc_feature 55855. .58174
/note="assembly_fragment:00549"
misc_feature 58275. .60640
/note="assembly_fragment:00556"
misc_feature 60741. .64375
/note="assembly_fragment:00581"
misc_feature 64476. .67814
/note="assembly_fragment:00589"
misc_feature 67915. .70067
/note="assembly_fragment:00718"
misc_feature 70168. .72968
/note="assembly_fragment:00889"
misc_feature 73069. .75340
/note="assembly_fragment:00949"
misc_feature 75441. .78874
/note="assembly_fragment:01032"
BASE COUNT 21338 a 16690 c 16975 g 21759 t 2112 others
ORIGIN

```

```

Query Match      8.0%;  Score 35.4;  DB 2;  Length 78874;
Best Local Similarity 50.3%;  Pred. No. 9;
Matches 87;  Conservative 0;  Mismatches 86;  Indels 0;  Gaps 0;

```

Qy 61 ataatccgggtgccttccaagaatcctccaaccacccttggtgcccaagcaagccacaaa 120  
 || | | |||| || | | | | | | | | | | | | | | | | |  
 Db 18521 ATGCCCTGTGTGCTTTGGAACACGTGCACAACCACACCTTGTTTCATCACCATCCCAGAAA 18580

Qy 121 aagtgggcaaaccaaagaggacagcaatgctaggaaaatgacgatgacaaagacgagggc 180  
 || | | |||| || | | | | | | | | | | | | | | | | |  
 Db 18581 CCCTGACGCAGGCAAAGAGCAGAGTTATTAACCTACTTTACTGATGTGGATACTGAGGC 18640

Qy 181 atcgggcaacatacttggttagccgtaatgacgacggggccatgctatctagatt 233  
 | | ||| | |||| | |||| | | | | | | | | | | | | | |  
 Db 18641 CCAGAGGCTCATGCAAGTTATCAGTAAGTGGCAGGGACAGTTGCCTCTAGATT 18693

RESULT 15

AC064821

LOCUS AC064821 174986 bp DNA HTG 07-JUL-2000

DEFINITION Homo sapiens chromosome 12 clone RP11-125G9, WORKING DRAFT  
 SEQUENCE, 9 unordered pieces.

ACCESSION AC064821

VERSION AC064821.2 GI:7770020

KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 174986)

AUTHORS Waterston,R.H.

TITLE The sequence of Homo sapiens clone

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 174986)

AUTHORS Waterston,R.H.

TITLE Direct Submission

JOURNAL Submitted (22-APR-2000) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA

COMMENT On May 11, 2000 this sequence version replaced gi:7637335.

----- Genome Center -----

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site:<http://genome.wustl.edu/gsc/index.shtml>

----- Project Information -----

Center project name: H\_NH0125G09

----- Summary Statistics -----

Sequencing vector: M13; 100%

Sequencing vector: plasmid; 0%

Chemistry: Dye-primer ET; 100% of reads

Chemistry: Dye-terminator Big Dye; 0% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 170476 bases at least Q40

Consensus quality: 171216 bases at least Q30

Consensus quality: 171956 bases at least Q20

Insert size: 189000; agarose-fp

Insert size: 174186; sum-of-contigs

Quality coverage: 5.78 in Q20 bases; agarose-fp

Quality coverage: 6.32 in Q20 bases; sum-of-contigs

-----



\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 9 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1 2309: contig of 2309 bp in length  
 \* 2310 2409: gap of unknown length  
 \* 2410 5742: contig of 3333 bp in length  
 \* 5743 5842: gap of unknown length  
 \* 5843 12706: contig of 6864 bp in length  
 \* 12707 12806: gap of unknown length  
 \* 12807 23411: contig of 10605 bp in length  
 \* 23412 23511: gap of unknown length  
 \* 23512 38025: contig of 14514 bp in length  
 \* 38026 38125: gap of unknown length  
 \* 38126 55106: contig of 16981 bp in length  
 \* 55107 55206: gap of unknown length  
 \* 55207 85998: contig of 30792 bp in length  
 \* 85999 86098: gap of unknown length  
 \* 86099 124167: contig of 38069 bp in length  
 \* 124168 124267: gap of unknown length  
 \* 124268 174986: contig of 50719 bp in length.

FEATURES	Location/Qualifiers
source	1. .174986 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="12" /clone="RP11-125G9"
misc_feature	1. .2309 /note="assembly_name:Contig2"
misc_feature	2410. .5742 /note="assembly_name:Contig3"
misc_feature	5843. .12706 /note="assembly_name:Contig4"
misc_feature	12807. .23411 /note="assembly_name:Contig5"
misc_feature	23512. .38025 /note="assembly_name:Contig6"
misc_feature	38126. .55106 /note="assembly_name:Contig7" clone_end:T7 vector_side:right"
misc_feature	55207. .85998 /note="assembly_name:Contig8"
misc_feature	86099. .124167 /note="assembly_name:Contig9"
misc_feature	124268. .174986 /note="assembly_name:Contig10" clone_end:SP6 vector_side:right"
BASE COUNT	55960 a 31805 c 33275 g 53145 t 801 others
ORIGIN	

Query Match 8.0%; Score 35.4; DB 2; Length 174986;  
Best Local Similarity 51.6%; Pred. No. 9.8;  
Matches 81; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

```
Qy 207 atgacgacgggccatgctatctagattccggtcttaatgagtacgtctgcagaaagacta 266
      ||| ||||| ||| ||| | | |||| | || | | |||| |||||
Db 70184 ATAAAGACGAGGATGATATTTGTTCTAGAATCTTGAAAAGAAGGAATGCAAAAAGAGGC 70243

Qy 267 ataagtgcataagagcttggtgctctgcgtggcgagttgtcaaccatcatcatgaattc 326
      | ||| ||| | | | || |||| | | |||| | | || ||
Db 70244 AAAAGAAGTAAATTAAGATTATGTTTCAGCGAAGTCATTTGTCTGGAAATTTATGTAACTC 70303

Qy 327 aagatactgcggagacatcatgatactgcggagacag 363
      || | | || | |||| | ||| |||
Db 70304 AAATAATTAGGGCAAACCTGTGATGTTTGGGATAAAG 70340
```

Search completed: February 7, 2002, 10:57:42  
Job time: 9388 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 10:59:37 ; Search time 428.31 Seconds  
(without alignments)  
888.731 Million cell updates/sec

Title: US-09-394-745-6154  
Perfect score: 444  
Sequence: 1 cgaaaacactgggtacccaaa.....tcccattttaagaaataaat 444

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_1101:\*

- 1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:\*
- 2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:\*
- 3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:\*
- 4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:\*
- 5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:\*
- 6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:\*
- 7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:\*

8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result	%		Query			ID	Description
	No.	Score	Match	Length	DB		
	1	284	64.0	379	20	AAX90965	Maize basal endosp
	2	35.2	7.9	1920	19	AAV17563	Coding sequence fo
c	3	34.2	7.7	1404	13	AAQ25079	Alpha-amylase vari
c	4	34.2	7.7	1404	15	AAQ77667	Variant alpha amyl
c	5	34.2	7.7	1404	15	AAQ77668	Variant alpha amyl
c	6	34.2	7.7	1404	15	AAQ77665	Variant alpha amyl
c	7	34.2	7.7	1404	15	AAQ77666	Variant alpha amyl
c	8	34.2	7.7	4214	8	AAN70916	Sequence encoding
	9	33.8	7.6	96109	22	AAF28548	Genomic fragment #
c	10	33.4	7.5	1120	21	AAC46213	Arabidopsis thalia
c	11	33.4	7.5	1123	21	AAC35071	Arabidopsis thalia
c	12	32.4	7.3	273	22	AAI24374	Probe #14307 for g
c	13	32.4	7.3	273	22	AAI09913	Probe #9904 used t
c	14	31.8	7.2	7607	14	AAQ49754	pTK gene LpTK-2.
c	15	31.8	7.2	7607	16	AAT03097	Protein tyrosine-k
c	16	31.6	7.1	323	13	AAQ30992	Notch clone hN4k s
	17	31	7.0	3542	21	AAA53893	S-adenosylmethioni
c	18	31	7.0	183999	22	AAF92831	Human ABC1 genomic
	19	30.8	6.9	2277	19	AAV13834	Homo sapiens ambig
	20	30.8	6.9	2277	19	AAV05370	Human telomerase p
	21	30.8	6.9	4338	21	AAA57197	Human subtilisin-k
c	22	30.2	6.8	477	22	AAI12071	Probe #2004 for ge
c	23	30.2	6.8	477	22	AAI33408	Probe #2094 used t
c	24	30.2	6.8	477	22	AAI01994	Probe #1985 used t
c	25	30.2	6.8	700	22	AAH92225	Human inflammatory
	26	30.2	6.8	25871	21	AAA09888	Human genomic OCTN
c	27	30.2	6.8	162450	21	AAZ86967	Retinoblastoma bin
	28	30	6.8	521	22	AAI41607	Probe #10293 used
	29	29.8	6.7	765	21	AAF14350	Aspergillus oryzae
	30	29.8	6.7	2295	17	AAT31994	Nonsense-mediated
	31	29.8	6.7	2295	21	AAA39451	Yeast NMD2 carboxy
	32	29.8	6.7	4080	17	AAT31993	Nonsense-mediated

	33	29.8	6.7	4080	21	AAA39450	Yeast NMD2 gene.
	34	29.8	6.7	4082	20	AAX25601	Yeast NMD2 gene in
	35	29.4	6.6	1968	22	AAH15153	Human cDNA sequenc
	36	29.4	6.6	7900	20	AAX13068	Enterococcus faeca
	37	29.2	6.6	1431	22	AAF27861	Human NOV11 cDNA.
	38	29.2	6.6	1743	22	AAS01213	DNA encoding human
c	39	29.2	6.6	2884	21	AAA79490	Eucalyptus grandis
c	40	29.2	6.6	3600	21	AAA79657	Eucalyptus grandis
	41	29	6.5	1212	20	AAZ31543	Generic mouse cycl
	42	29	6.5	1212	22	AAD02493	Mouse cyclin E2 am
	43	28.8	6.5	489	21	AAC94433	Cat flea hindgut a
c	44	28.8	6.5	936	22	AAF58252	Oligonucleotide D1
c	45	28.8	6.5	936	22	AAF58254	Oligonucleotide D1

# ALIGNMENTS

## RESULT 1

AAX90965

ID AAX90965 standard; cDNA; 379 BP.

XX

AC AAX90965;

XX

DT 17-JAN-2000 (first entry)

XX

DE Maize basal endosperm transfer cell layer-2 cDNA.

XX

KW Maize basal endosperm transfer cell layer-2 specific protein; BETL-2;  
 KW defensin supergene family; antimicrobial peptide; endosperm; promoter;  
 KW grain development; regulatory element; transgenic plant;  
 KW protein expression; BETL-specific expression; heterologous DNA;  
 KW solute partitioning; disease resistance; endosperm-derived product;  
 KW cotton quality; aromatic oil; ss.

XX

OS Zea mays.

XX

FH Key Location/Qualifiers

FT CDS 44..328

FT /\*tag= a

FT /product= "Basal endosperm transfer cell layer-2 protein"

XX

PN WO9950427-A2.

XX

PD 07-OCT-1999.

XX

PF 26-MAR-1999; 99WO-EP02063.

XX

PR 27-MAR-1998; 98EP-0105628.

XX

PA (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

XX

PI Thompson RD, Yan G, Salamini F, Hueros G;

XX

DR WPI; 1999-610858/52.

DR P-PSDB; AAY28847.

XX

PT New nucleic acid encoding three basal endosperm transfer cell layer  
PT proteins, used to produce transgenic plants with e.g. increased disease  
PT resistance and to identify specific modulators -  
XX  
PS Claim 1; Page 64-65; 76pp; English.  
XX  
CC The present sequence encodes for the maize basal endosperm transfer cell  
CC layer-2 specific protein. This has homology to defensin supergene family  
CC of antimicrobial peptides. The basal region of endosperm is highly  
CC specialised to facilitate uptake of solutes during grain development.  
CC Vectors comprising this nucleic acid sequence operably linked to  
CC regulatory elements is used to produce transgenic plants. These plants  
CC have altered levels of BETL protein expression. The regulatory region of  
CC the promoter is used to provide BETL-specific expression of heterologous  
CC DNA; to modify solute partitioning in the endosperm; for disease  
CC resistance; to improve endosperm-derived products and to express enzymes  
CC that modify quality of cotton or aromatic oils.  
XX  
SQ Sequence 379 BP; 107 A; 85 C; 94 G; 93 T; 0 other;

Query Match 64.0%; Score 284; DB 20; Length 379;  
Best Local Similarity 87.5%; Pred. No. 1.7e-84;  
Matches 322; Conservative 0; Mismatches 45; Indels 1; Gaps 1;

```

Qy      3 aaaacactggtacccaaaacaaccgtcaaccaagggcaaattcaacaacctccaaagaat 62
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     13 agactattgtagctcatatcatctgtcacccatggcgaagtgcagcagcttccaaggatt 72

Qy     63 aatccgggtgccttccaagaatcctccaaccacccttggtgcccaagcaagccacaaaaa 122
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     73 attctggttgctttccatgattcttctagcatcctttggtgctcatgcacg-cacaaca 131

Qy    123 gtgggcaaaccaaagaggacagcaatgctaggaaaatgacgatgacaaagacgagggcat 182
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    132 gtgggcaaaccaaagaggacagcaatgctaggaacatgacgatgaccaagacgagggcat 191

Qy    183 cgggcaacatacttgtagccgtaatgacgacgggccatgctatctagattccgggtctta 242
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    192 caggcaacatacttgtagccgtaatgacgacgggccatgctatctagattccgggtctta 251

Qy    243 atgagtacgtctgcagaaagactaataagtgtctataagagcttggtgctctgcgtggcga 302
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    252 atgagtacgtctgcagaaagactaataagtgtctataagagcttggtgctctgcgtggcga 311

Qy    303 gttgtcaaccatcatcatgaattcaagatactgcggagacatcatgatactgcggagaca 362
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    312 gttgtcaaccatcatcatgaattcatgatactgcggagacatcatgatactgcggagaca 371

Qy    363 gacggcca 370
      | | | | | |
Db    372 gacggcga 379

```

RESULT 2  
AAV17563  
ID AAV17563 standard; cDNA; 1920 BP.

XX  
 AC AAV17563;  
 XX  
 DT 10-JUN-1998 (first entry)  
 XX  
 DE Coding sequence for the alpha' subunit of beta-conglycinin.  
 XX  
 KW Beta-conglycinin; soybean seed protein; transgenic plant;  
 KW seed storage protein profile; ss.  
 XX  
 OS Glycine max.  
 XX  
 PN WO9747731-A2.  
 XX  
 PD 18-DEC-1997.  
 XX  
 PF 10-JUN-1997; 97WO-US09743.  
 XX  
 PR 14-JUN-1996; 96US-0019940.  
 XX  
 PA (DUPO ) DU PONT DE NEMOURS & CO E I.  
 XX  
 PI Fader GM, Kinney AJ;  
 XX  
 DR WPI; 1998-052298/05.  
 XX  
 PT Suppression of specific classes of soybean seed protein genes -  
 PT useful to change seed storage protein profiles of transgenic plants  
 XX  
 PS Disclosure; Page 30-31; 58pp; English.  
 XX  
 CC This sequence represents the coding sequence for the alpha' subunit of  
 CC the soybean seed protein beta-conglycinin. The method of the invention is  
 CC for reducing the quantity of a soybean seed storage protein (A), such as  
 CC beta-conglycinin, in soybeans. It comprises: (a) constructing a chimeric  
 CC gene comprising: (i) a nucleic acid fragment encoding a promoter that is  
 CC functional in the cells of soybean seeds; (ii) a nucleic acid fragment  
 CC encoding all or a portion of (A) placed in sense or antisense orientation  
 CC relative to the promoter of (i); and (iii) a transcriptional termination  
 CC region; (b) creating a transgenic soybean cell by introducing into a  
 CC soybean cell the chimeric gene of (a); and (c) growing the transgenic  
 CC soybean cells of (b) under conditions that result in expression of the  
 CC chimeric gene of (a); where the quantity of one or more members of a  
 CC class of (A) subunits is reduced when compared to soybeans not containing  
 CC the chimeric gene of (a). The method is used to construct transgenic  
 CC soybean lines where the expression of genes encoding (A) are modulated to  
 CC effect a change in seed storage protein profile of transgenic plants.  
 CC Modification of the seed storage protein profile can result in the  
 CC production of novel soy protein products with unique and valuable  
 CC functional characteristics.  
 XX  
 SQ Sequence 1920 BP; 634 A; 444 C; 449 G; 393 T; 0 other;

Query Match 7.9%; Score 35.2; DB 19; Length 1920;  
 Best Local Similarity 62.5%; Pred. No. 0.18;  
 Matches 55; Conservative 0; Mismatches 33; Indels 0; Gaps 0;



XX

SQ Sequence 1404 BP; 407 A; 271 C; 294 G; 432 T; 0 other;

Query Match 7.7%; Score 34.2; DB 13; Length 1404;  
Best Local Similarity 49.7%; Pred. No. 0.33;  
Matches 87; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

```
Qy 51 cctccaaagaataatccgggtgccttccaagaatcctccaaccacccttggtgcccaagc 110
    ||| | || | | || | | || | | || | | || | | || | | || | |
Db 689 CCTACTGAGTAAACTCCAGATGCACTAACAAAATCCGGGAAAAAGCCTTGGTCCACATGT 630

Qy 111 aagccacaaaaagtgggcaaaccaaagaggacagcaatgctaggaaaatgacgatgacaa 170
    ||| | | | | || | | | | | | | | | | | | | | | |
Db 629 TTAGCACTATCAATTCTTAAACCATCAATTGAGTAATTGCCAACAAAATCTTTAACCCAA 570

Qy 171 agacgagggcatcgggcaacatacttgtagccgtaatgacgacgggccaatgcta 225
    | | | | | | | | | | | | | | | | | | | | | |
Db 569 GAATTGAAAACCTGAGGCCACCTCGCTATCTTCGTTCTCAAATCTGGTAATGCAA 515
```

RESULT 4

AAQ77667/c

ID AAQ77667 standard; DNA; 1404 BP.

XX

AC AAQ77667;

XX

DT 16-JUN-1995 (first entry)

XX

DE Variant alpha amylase deriv. from *Saccharomyopsis fibuligera*.

XX

KW alpha amylase; carbohydrate hydrolase; increased activity;  
KW tyrosine residue; enzyme centre; mass production; oligosaccharide;  
KW variant; cyclomaltodextrin glucanotransferase; ds.

XX

OS *Saccharomycopsis fibuligera*.

XX

FH Key Location/Qualifiers

FT misc\_difference 247..249

FT /\*tag= a

FT /note= "the wild type sequence TAY was mutated to  
FT CTC to give a variant enzyme"

XX

PN JP06253836-A.

XX

PD 13-SEP-1994.

XX

PF 04-MAR-1993; 93JP-0069303.

XX

PR 04-MAR-1993; 93JP-0069303.

XX

PA (AGEN ) AGENCY OF IND SCI & TECHNOLOGY.

XX

DR WPI; 1994-328987/41.

DR P-PSDB; AAR63186.

XX

PT Variant carbohydrate hydrolase(s) with increased activity -



PT consists of e.g. alpha-amylase with tyrosine residue in enzyme  
 PT centre, useful for mass-prodn. of oligosaccharide(s)  
 XX  
 PS Example 1; Page 18-20; 27pp; Japanese.  
 XX  
 CC AAQ77665-8 encode variant alpha amylases, composed by substituting  
 CC bases 247-249 of the structural gene region, with TTC, TGG, CTC or  
 CC AAC. These substitutions result in the 83rd amino acid residue  
 CC (tyrosine) of the wild type sequence being changed to phenylalanine,  
 CC tryptophan, leucine or asparagine respectively. The substituted  
 CC amino acid is present in the active site of the enzyme and confers  
 CC increased activity on the enzyme. The variants are useful for the  
 CC mass production of oligosaccharides. (see AAQ77669 for the variant  
 CC structure of a cyclomaltodextrin glucanotransferase).  
 XX  
 SQ Sequence 1404 BP; 406 A; 273 C; 293 G; 432 T; 0 other;

Query Match 7.7%; Score 34.2; DB 15; Length 1404;  
 Best Local Similarity 49.7%; Pred. No. 0.33;  
 Matches 87; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

```

Qy      51 cctccaaagaataatccgggtgccttccaagaatcctccaaccacccttggtgcccaagc 110
      ||| | || | | ||| | ||| | | |||| | | | ||||| | || |
Db      689 CCTACTGAGTAAACTCCAGATGCACTAACAAAATCCGGGAAAAAGCCTTGGTCCACATGT 630

Qy     111 aagccacaaaaagtgggcaaaccaaagaggacagcaatgctaggaaaatgacgatgacaa 170
      ||| | | | | ||||| | | ||| | |||| | | |||
Db     629 TTAGCACTATCAATTCTTAAACCATCAATTGAGTAATTGCCAACAAAATCTTTAACCCAA 570

Qy     171 agacgagggcatcgggcaacatacttgtagccgtaatgacgacgggcatgcta 225
      | | | ||| || | | | |||| | | | ||| |||| |
Db     569 GAATTGAAAACCTGAGGCCACGTCGCTATCTTCCGTTCTCAAATCTGGTAATGCAA 515

```

RESULT 5  
 AAQ77668/c  
 ID AAQ77668 standard; DNA; 1404 BP.  
 XX  
 AC AAQ77668;  
 XX  
 DT 16-JUN-1995 (first entry)  
 XX  
 DE Variant alpha amylase deriv. from *Saccharomyopsis fibuligera*.  
 XX  
 KW alpha amylase; carbohydrate hydrolase; increased activity;  
 KW tyrosine residue; enzyme centre; mass production; oligosaccharide;  
 KW variant; cyclomaltodextrin glucanotransferase; ds.  
 XX  
 OS *Saccharomycopsis fibuligera*.  
 XX  
 FH Key Location/Qualifiers  
 FT misc\_difference 247..249  
 FT /\*tag= a  
 FT /note= "the wild type sequence TAY was mutated to  
 FT AAC to give a variant enzyme"  
 XX

PN JP06253836-A.  
 XX  
 PD 13-SEP-1994.  
 XX  
 PF 04-MAR-1993; 93JP-0069303.  
 XX  
 PR 04-MAR-1993; 93JP-0069303.  
 XX  
 PA (AGEN ) AGENCY OF IND SCI & TECHNOLOGY.  
 XX  
 DR WPI; 1994-328987/41.  
 DR P-PSDB; AAR63187.  
 XX  
 PT Variant carbohydrate hydrolase(s) with increased activity -  
 PT consists of e.g. alpha-amylase with tyrosine residue in enzyme  
 PT centre, useful for mass-prodn. of oligosaccharide(s)  
 XX  
 PS Example 1; Page 20-23; 27pp; Japanese.  
 XX  
 CC AAQ77665-8 encode variant alpha amylases, composed by substituting  
 CC bases 247-249 of the structural gene region, with TTC, TGG, CTC or  
 CC AAC. These substitutions result in the 83rd amino acid residue  
 CC (tyrosine) of the wild type sequence being changed to phenylalanine,  
 CC tryptophan, leucine or asparagine respectively. The substituted  
 CC amino acid is present in the active site of the enzyme and confers  
 CC increased activity on the enzyme. The variants are useful for the  
 CC mass production of oligosaccharides. (see AAQ77669 for the variant  
 CC structure of a cyclomaltodextrin glucanotransferase).  
 XX  
 SQ Sequence 1404 BP; 408 A; 272 C; 293 G; 431 T; 0 other;

Query Match 7.7%; Score 34.2; DB 15; Length 1404;  
 Best Local Similarity 49.7%; Pred. No. 0.33;  
 Matches 87; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

Qy 51 cctccaaagaataatccgggtgccttccaagaatcctccaaccacccttggtgcccaagc 110  
 ||| | || | ||| | ||| | | |||| | | | ||||| | || |  
 Db 689 CCTACTGAGTAAACTCCAGATGCACTAACAAAATCCGGGAAAAAGCCTTGGTCCACATGT 630  
 Qy 111 aagccacaaaaagtgggcaaaccaaagaggacagcaatgctaggaaaatgacgatgacaa 170  
 ||| | | | ||||| | | ||| | |||| | | |||  
 Db 629 TTAGCACTATCAATTCTTAAACCATCAATTGAGTAATTGCCAACAAAATCTTTAACCCAA 570  
 Qy 171 agacgagggcatcgggcaacatacttggttagccgtaatgacgacgggcatgcta 225  
 | | ||| || | | |||| | | | ||| |||| |  
 Db 569 GAATTGAAAACCTGAGGCCACGTCGCTATCTTCCGTTCTCAAATCTGGTAATGCAA 515

RESULT 6  
 AAQ77665/c  
 ID AAQ77665 standard; DNA; 1404 BP.  
 XX  
 AC AAQ77665;  
 XX  
 DT 16-JUN-1995 (first entry)  
 XX

DE Variant alpha amylase deriv. from Saccharomyopsis fibuligera.  
 XX  
 KW alpha amylase; carbohydrate hydrolase; increased activity;  
 KW tyrosine residue; enzyme centre; mass production; oligosaccharide;  
 KW variant; cyclomaltodextrin glucanotransferase; ds.  
 XX  
 OS Saccharomycopsis fibuligera.  
 XX  
 FH Key Location/Qualifiers  
 FT misc\_difference 247..249  
 FT /\*tag= a  
 FT /note= "the wild type sequence TAY was mutated to  
 FT TTC to give a variant enzyme"  
 XX  
 PN JP06253836-A.  
 XX  
 PD 13-SEP-1994.  
 XX  
 PF 04-MAR-1993; 93JP-0069303.  
 XX  
 PR 04-MAR-1993; 93JP-0069303.  
 XX  
 PA (AGEN ) AGENCY OF IND SCI & TECHNOLOGY.  
 XX  
 DR WPI; 1994-328987/41.  
 DR P-PSDB; AAR63184.  
 XX  
 PT Variant carbohydrate hydrolase(s) with increased activity -  
 PT consists of e.g. alpha-amylase with tyrosine residue in enzyme  
 PT centre, useful for mass-prodn. of oligosaccharide(s)  
 XX  
 PS Example 1; Page 13-15; 27pp; Japanese.  
 XX  
 CC AAQ77665-8 encode variant alpha amylases, composed by substituting  
 CC bases 247-249 of the structural gene region, with TTC, TGG, CTC or  
 CC AAC. These substitutions result in the 83rd amino acid residue  
 CC (tyrosine) of the wild type sequence being changed to phenylalanine,  
 CC tryptophan, leucine or asparagine respectively. The substituted  
 CC amino acid is present in the active site of the enzyme and confers  
 CC increased activity on the enzyme. The variants are useful for the  
 CC mass production of oligosaccharides. (see AAQ77669 for the variant  
 CC structure of a cyclomaltodextrin glucanotransferase).  
 XX  
 SQ Sequence 1404 BP; 406 A; 272 C; 293 G; 433 T; 0 other;

Query Match 7.7%; Score 34.2; DB 15; Length 1404;  
 Best Local Similarity 49.7%; Pred. No. 0.33;  
 Matches 87; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

Qy 51 cctccaaagaataatccgggtgccttccaagaatcctccaaccacccttggtgcccaagc 110  
 ||| | || | || | || | || | || | || | || | || | || | || | || |  
 Db 689 CCTACTGAGTAAACTCCAGATGCACTAACAAAATCCGGGAAAAAGCCTTGGTCCACATGT 630  
 Qy 111 aagccacaaaaagtgggcaaaccaaagaggacagcaatgctaggaaaatgacgatgacaa 170  
 ||| | | | |||| | | || | |||| | || |  
 Db 629 TTAGCACTATCAATTCTTAAACCATCAATTGAGTAATTGCCAACAAAATCTTTAACCCAA 570

Qy 171 agacgagggcatcgggcaacatacttggttagccgtaatgacgacgggccatgcta 225  
 | | | | | | | | | | | | | | | |  
 Db 569 GAATTGAAACTGAGGCCACGTCGCTATCTCCGTTCTCAAATCTGGTAATGCAA 515

RESULT 7

AAQ77666/c

ID AAQ77666 standard; DNA; 1404 BP.

XX

AC AAQ77666;

XX

DT 16-JUN-1995 (first entry)

XX

DE Variant alpha amylase deriv. from *Saccharomyopsis fibuligera*.

XX

KW alpha amylase; carbohydrate hydrolase; increased activity;

KW tyrosine residue; enzyme centre; mass production; oligosaccharide;

KW variant; cyclomaltodextrin glucanotransferase; ds.

XX

OS *Saccharomycopsis fibuligera*.

XX

FH Key Location/Qualifiers

FT misc\_difference 247..249

FT /\*tag= a

FT /note= "the wild type sequence TAY was mutated to

FT TGG to give a variant enzyme"

XX

PN JP06253836-A.

XX

PD 13-SEP-1994.

XX

PF 04-MAR-1993; 93JP-0069303.

XX

PR 04-MAR-1993; 93JP-0069303.

XX

PA (AGEN ) AGENCY OF IND SCI & TECHNOLOGY.

XX

DR WPI; 1994-328987/41.

DR P-PSDB; AAR63185.

XX

PT Variant carbohydrate hydrolase(s) with increased activity -

PT consists of e.g. alpha-amylase with tyrosine residue in enzyme

PT centre, useful for mass-prodn. of oligosaccharide(s)

XX

PS Example 1; Page 15-17; 27pp; Japanese.

XX

CC AAQ77665-8 encode variant alpha amylases, composed by substituting  
 CC bases 247-249 of the structural gene region, with TTC, TGG, CTC or  
 CC AAC. These substitutions result in the 83rd amino acid residue  
 CC (tyrosine) of the wild type sequence being changed to phenylalanine,  
 CC tryptophan, leucine or asparagine respectively. The substituted  
 CC amino acid is present in the active site of the enzyme and confers  
 CC increased activity on the enzyme. The variants are useful for the  
 CC mass production of oligosaccharides. (see AAQ77669 for the variant  
 CC structure of a cyclomaltodextrin glucanotransferase).

XX

SQ Sequence 1404 BP; 406 A; 271 C; 295 G; 432 T; 0 other;

Query Match 7.7%; Score 34.2; DB 15; Length 1404;  
Best Local Similarity 49.7%; Pred. No. 0.33;  
Matches 87; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

```
Qy 51 cctccaaagaataatccgggtgccttccaagaatcctccaaccacccttggtgcccaagc 110
    ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 689 CCTACTGAGTAAACTCCAGATGCACTAACAAAATCCGGGAAAAAGCCTTGGTCCACATGT 630

Qy 111 aagccacaaaaagtgggcaaaccaaagaggacagcaatgctaggaaaatgacgatgacaa 170
    ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 629 TTAGCACTATCAATTCTTAAACCATCAATTGAGTAATTGCCAACAAAATCTTTAACCCAA 570

Qy 171 agacgagggcatcgggcaacatacttggttagccgtaatgacgacgggcatgcta 225
    | | | | | | | | | | | | | | | | | | | | | |
Db 569 GAATTGAAAAC TGAGGCCACGTCGCTATCTTCCGTTCTCAAATCTGGTAATGCAA 515
```

RESULT 8

AAN70916/c

ID AAN70916 standard; DNA; 4214 BP.

XX

AC AAN70916;

XX

DT 03-MAY-1991 (first entry)

XX

DE Sequence encoding alpha-amylase from plasmid pSf alpha 1.

XX

KW Amylase; ds.

XX

OS *Saccharomyces fibuligera* HUT7212.

XX

FH Key Location/Qualifiers

FT CDS 1531..3015

FT /\*tag= a

XX

PN JP62104576-A.

XX

PD 15-MAY-1987.

XX

PF 31-OCT-1985; 85JP-0244892.

XX

PR 31-OCT-1985; 85JP-0244892.

XX

PA (FUKU/) FUKUI S.

XX

DR WPI; 1987-173694/25.

DR P-PSDB; AAP70571.

XX

PT Amylase prodn. - comprises culturing microorganism transformed  
PT with vector deoxyribonucleic acid, accumulating and collecting  
PT amylase

XX

PS Disclosure; Fig 1; 14pp; Japanese.

XX

CC The plasmid may be used to transform an E.coli expression system for  
CC the stable production of amylase, useful in ethanol fermentation.  
CC See also AAN70917.  
XX  
SQ Sequence 4214 BP; 1249 A; 784 C; 860 G; 1321 T; 0 other;

Query Match 7.7%; Score 34.2; DB 8; Length 4214;  
Best Local Similarity 49.7%; Pred. No. 0.55;  
Matches 87; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

Qy 51 cctccaaagaataatccgggtgccttccaagaatcctccaaccacccttggtgcccaagc 110  
||| | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 2297 CCTACTGAGTAAACTCCAGATGCACTAACAAAATCCGGGAAAAAGCCTTGGTCCACATGT 2238  
  
Qy 111 aagccacaaaaagtgggcaaaccaaagaggacagcaatgctaggaaaatgacgatgacaa 170  
||| | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 2237 TTAGCACTATCAATTCTTAAACCATCAATTGAGTAATTGCCAACAAAATCTTTAACCCAA 2178  
  
Qy 171 agacgagggcatcgggcaacatacttggttagccgtaatgacgacgggccaatgcta 225  
| | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 2177 GAATTGAAAACCTGAGGCCACGTCGCTATCTCCGTTCTCAAATCTGGTAATGCAA 2123

RESULT 9

AAF28548

ID AAF28548 standard; DNA; 96109 BP.

XX

AC AAF28548;

XX

DT 04-APR-2001 (first entry)

XX

DE Genomic fragment #35.

XX

KW Genomic library; bacteria; human upper airway; otitis media; sinusitis;

KW bronchopulmonary; endocarditis; meningitis; ss.

XX

OS Moraxella catarrhalis.

XX

PN WO200078968-A2.

XX

PD 28-DEC-2000.

XX

PF 16-JUN-2000; 2000WO-US16649.

XX

PR 18-JUN-1999; 99US-0140121.

XX

PA (INCY-) INCYTE GENOMICS INC.

XX

PI Lagace RE, Patterson C, Berg KL;

XX

DR WPI; 2001-041427/05.

XX

PT Genomic library for identifying diagnostic and therapeutic

PT compositions, and for identifying virulence factors, regulatory

PT elements and drug targets, comprises Moraxella catarrhalis nucleic

PT acids -

XX  
 PS Claim 1; Page 345-368; 545pp; English.  
 XX  
 CC The present invention relates to a *Moraxella catarrhalis* genomic library  
 CC comprising of a combination of 41 nucleic acid molecules (see  
 CC AAF28514-AAF28554). The library has a number of uses described in the  
 CC specification e.g. is useful for identifying diagnostic and therapeutic  
 CC compositions. *M. catarrhalis* (*Branhamella catarrhalis*) is a large  
 CC aerobic, gram-negative diplococcus, normally found among the bacterial  
 CC flora of human upper airways. *M. catarrhalis* is known to cause acute,  
 CC localised infections such as otitis media, sinusitis and bronchopulmonary  
 CC infection and life-threatening, systemic diseases including endocarditis  
 CC and meningitis.  
 XX  
 SQ Sequence 96109 BP; 28783 A; 18910 C; 20341 G; 28075 T; 0 other;

Query Match 7.6%; Score 33.8; DB 22; Length 96109;  
 Best Local Similarity 53.4%; Pred. No. 3.2;  
 Matches 71; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy 2 gaaaacactgggtacccaaaacaacggtcaaccaagggcaaattcaacaacctccaaagaa 61  
 |||| | || | |||| ||| | | |||| || || ||| | ||| | ||  
 Db 80672 gaaatcgctaattgcccatcccaaacctgacccaataccatatacaatagactcaccaaaa 80731  
 Qy 62 taatccgggtgccttccaagaatcctccaaccacccttggtgcccaagcaagccacaaaa 121  
 | || | || | || || | |||| || || | || | || | ||  
 Db 80732 ttataatcacgctcaaccataaacaatacaccacccatgatggcacagtctcactgtaatc 80791  
 Qy 122 agtgggcaaacca 134  
 | || ||| ||  
 Db 80792 aagggtaaaaaaca 80804

RESULT 10  
 AAC46213/c  
 ID AAC46213 standard; DNA; 1120 BP.  
 XX  
 AC AAC46213;  
 XX  
 DT 18-OCT-2000 (first entry)  
 XX  
 DE Arabidopsis thaliana DNA fragment SEQ ID NO: 49314.  
 XX  
 KW Hybridisation assay; genetic mapping; gene expression control;  
 KW protein identification; signal transduction pathway;  
 KW metabolic pathway; promoter; termination sequence; ss.  
 XX  
 OS Arabidopsis thaliana.  
 XX  
 PN EP1033405-A2.  
 XX  
 PD 06-SEP-2000.  
 XX  
 PF 25-FEB-2000; 2000EP-0301439.  
 XX  
 PR 25-FEB-1999; 99US-0121825.

PR	05-MAR-1999;	99US-0123180.
PR	09-MAR-1999;	99US-0123548.
PR	23-MAR-1999;	99US-0125788.
PR	25-MAR-1999;	99US-0126264.
PR	29-MAR-1999;	99US-0126785.
PR	01-APR-1999;	99US-0127462.
PR	06-APR-1999;	99US-0128234.
PR	08-APR-1999;	99US-0128714.
PR	16-APR-1999;	99US-0129845.
PR	19-APR-1999;	99US-0130077.
PR	21-APR-1999;	99US-0130449.
PR	23-APR-1999;	99US-0130510.
PR	23-APR-1999;	99US-0130891.
PR	28-APR-1999;	99US-0131449.
PR	30-APR-1999;	99US-0132048.
PR	30-APR-1999;	99US-0132407.
PR	04-MAY-1999;	99US-0132484.
PR	05-MAY-1999;	99US-0132485.
PR	06-MAY-1999;	99US-0132486.
PR	06-MAY-1999;	99US-0132487.
PR	07-MAY-1999;	99US-0132863.
PR	11-MAY-1999;	99US-0134256.
PR	14-MAY-1999;	99US-0134218.
PR	14-MAY-1999;	99US-0134219.
PR	14-MAY-1999;	99US-0134221.
PR	14-MAY-1999;	99US-0134370.
PR	18-MAY-1999;	99US-0134768.
PR	19-MAY-1999;	99US-0134941.
PR	20-MAY-1999;	99US-0135124.
PR	21-MAY-1999;	99US-0135353.
PR	24-MAY-1999;	99US-0135629.
PR	25-MAY-1999;	99US-0136021.
PR	27-MAY-1999;	99US-0136392.
PR	28-MAY-1999;	99US-0136782.
PR	01-JUN-1999;	99US-0137222.
PR	03-JUN-1999;	99US-0137528.
PR	04-JUN-1999;	99US-0137502.
PR	07-JUN-1999;	99US-0137724.
PR	08-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	10-JUN-1999;	99US-0138847.
PR	14-JUN-1999;	99US-0139119.
PR	16-JUN-1999;	99US-0139452.
PR	16-JUN-1999;	99US-0139453.
PR	17-JUN-1999;	99US-0139492.
PR	18-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139458.
PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139750.
PR	18-JUN-1999;	99US-0139763.



PR	21-JUN-1999;	99US-0139817.
PR	22-JUN-1999;	99US-0139899.
PR	23-JUN-1999;	99US-0140353.
PR	23-JUN-1999;	99US-0140354.
PR	24-JUN-1999;	99US-0140695.
PR	28-JUN-1999;	99US-0140823.
PR	29-JUN-1999;	99US-0140991.
PR	30-JUN-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141842.
PR	01-JUL-1999;	99US-0142154.
PR	02-JUL-1999;	99US-0142055.
PR	06-JUL-1999;	99US-0142390.
PR	08-JUL-1999;	99US-0142803.
PR	09-JUL-1999;	99US-0142920.
PR	12-JUL-1999;	99US-0142977.
PR	13-JUL-1999;	99US-0143542.
PR	14-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	16-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.

PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.

PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 7.5%; Score 33.4; DB 21; Length 1120;  
 Best Local Similarity 52.2%; Pred. No. 0.55;  
 Matches 70; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

Qy 305 tgtcaaccatcatcatgaattcaagatactgCGGAGACATCATGATACTGCGGAGACAGA 364  
 | | | | | | | | | | | | | | | | | | | | | |  
 Db 767 TCTTCATCATCGTCATAATATTCATCACCTGCGCGGACAGAGTGAACTCAGGAGGTGGA 708  
 Qy 365 CGGCCAGAGATGANGCTAGCTAGATGCCGTTTCACCANNATATTATGTAACACCCAAATC 424  
 | | | | | | | | | | | | | | | | | | | | | |  
 Db 707 GGACAGTCTTCGGTTATGGCCTCATCACGTTTCATCACAAGATGCCTAATCACCTCTCA 648  
 Qy 425 tcccattttaagaa 438  
 | | | | | | | |  
 Db 647 TCACTATCTAACAA 634

# RESULT 11

AAC35071/c

ID AAC35071 standard; DNA; 1123 BP.

XX

AC AAC35071;

XX

DT 17-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 8891.

XX

KW Hybridisation assay; genetic mapping; gene expression control;

KW protein identification; signal transduction pathway;

KW metabolic pathway; promoter; termination sequence; ss.

XX

OS Arabidopsis thaliana.

XX

PN EP1033405-A2.

XX

PD 06-SEP-2000.

XX

PF 25-FEB-2000; 2000EP-0301439.

XX

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR	16-APR-1999;	99US-0129845.
PR	19-APR-1999;	99US-0130077.
PR	21-APR-1999;	99US-0130449.
PR	23-APR-1999;	99US-0130510.
PR	23-APR-1999;	99US-0130891.
PR	28-APR-1999;	99US-0131449.
PR	30-APR-1999;	99US-0132048.
PR	30-APR-1999;	99US-0132407.
PR	04-MAY-1999;	99US-0132484.
PR	05-MAY-1999;	99US-0132485.
PR	06-MAY-1999;	99US-0132486.
PR	06-MAY-1999;	99US-0132487.
PR	07-MAY-1999;	99US-0132863.
PR	11-MAY-1999;	99US-0134256.
PR	14-MAY-1999;	99US-0134218.
PR	14-MAY-1999;	99US-0134219.
PR	14-MAY-1999;	99US-0134221.
PR	14-MAY-1999;	99US-0134370.
PR	18-MAY-1999;	99US-0134768.
PR	19-MAY-1999;	99US-0134941.
PR	20-MAY-1999;	99US-0135124.
PR	21-MAY-1999;	99US-0135353.
PR	24-MAY-1999;	99US-0135629.
PR	25-MAY-1999;	99US-0136021.
PR	27-MAY-1999;	99US-0136392.
PR	28-MAY-1999;	99US-0136782.
PR	01-JUN-1999;	99US-0137222.
PR	03-JUN-1999;	99US-0137528.
PR	04-JUN-1999;	99US-0137502.
PR	07-JUN-1999;	99US-0137724.
PR	08-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	10-JUN-1999;	99US-0138847.
PR	14-JUN-1999;	99US-0139119.
PR	16-JUN-1999;	99US-0139452.
PR	16-JUN-1999;	99US-0139453.
PR	17-JUN-1999;	99US-0139492.
PR	18-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139458.
PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139750.
PR	18-JUN-1999;	99US-0139763.
PR	21-JUN-1999;	99US-0139817.
PR	22-JUN-1999;	99US-0139899.
PR	23-JUN-1999;	99US-0140353.
PR	23-JUN-1999;	99US-0140354.
PR	24-JUN-1999;	99US-0140695.
PR	28-JUN-1999;	99US-0140823.
PR	29-JUN-1999;	99US-0140991.
PR	30-JUN-1999;	99US-0141287.

PR	01-JUL-1999;	99US-0141842.
PR	01-JUL-1999;	99US-0142154.
PR	02-JUL-1999;	99US-0142055.
PR	06-JUL-1999;	99US-0142390.
PR	08-JUL-1999;	99US-0142803.
PR	09-JUL-1999;	99US-0142920.
PR	12-JUL-1999;	99US-0142977.
PR	13-JUL-1999;	99US-0143542.
PR	14-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	16-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.

PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

Query Match 7.5%; Score 33.4; DB 21; Length 1123;  
Best Local Similarity 52.2%; Pred. No. 0.55;  
Matches 70; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

```
Qy 305 tgtcaaccatcatcatgaattcaagatactgCGGagacatcatgatactgCGGagacaga 364
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 770 TCTTCATCATCGTCATAATATTCATCACCTGCGCGGACAGAGTGAAACTCAGGAGGTGGA 711

Qy 365 cggccagagatgangctagctagatgccgtttcaccannatattatgtaacacccaaatc 424
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 710 GGACAGTCTTCGGTTATGGCCTCATCAGTTTCATCACAAGATGCCTAATCACCTCTCA 651

Qy 425 tcccattttaagaa 438
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 650 TCACTATCTAACAA 637
```

RESULT 12  
AAI24374/c  
ID AAI24374 standard; DNA; 273 BP.  
XX  
AC AAI24374;  
XX  
DT 12-OCT-2001 (first entry)  
XX  
DE Probe #14307 for gene expression analysis in human cervical cell sample.  
XX  
KW Probe; human; microarray; gene expression; cervical epithelial cell;  
KW cervical cancer; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200157278-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00670.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
DR WPI; 2001-488901/53.  
XX  
PT Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human cervical epithelial cells -  
XX  
PS Claim 25; SEQ ID No 14307; 487pp; English.

XX  
CC The present invention relates to human single exon nucleic acid probes  
CC (SENP). The present sequence is one such probe. The SENPs are derived  
CC from human HeLa cells. The SENPs can be used to produce a single exon  
CC microarray, which can be used for measuring human gene expression in a  
CC sample derived from human cervical epithelial cells. By measuring gene  
CC expression, the probes are therefore useful in grading and/or staging  
CC of diseases of the cervix, notably cervical cancer.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 273 BP; 65 A; 54 C; 71 G; 83 T; 0 other;

Query Match 7.3%; Score 32.4; DB 22; Length 273;  
Best Local Similarity 56.6%; Pred. No. 0.62;  
Matches 60; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

Qy 71 tgccttccaagaatcctccaaccacccttggtgcccaagcaagccacaaaaagtgggcaa 130  
||||| ||| || ||| | | | | ||| | | || | ||  
Db 173 TGCCTTCCCCACCTCCCCCTCTTCCCCCTCCTCTACTACCAGCCCCCATAAACAGACAG 114

Qy 131 accaaagaggacagcaatgctaggaaaatgacgatgacaaagacga 176  
| | | ||| |||| | || | | | | | | | |  
Db 113 AAGACAAAGGAGTTCAATGTGAGGAAGAGGAAGAAGAGAAGAAAGA 68

RESULT 13  
AAI09913/c  
ID AAI09913 standard; DNA; 273 BP.  
XX  
AC AAI09913;  
XX  
DT 09-OCT-2001 (first entry)  
XX  
DE Probe #9904 used to measure gene expression in human breast sample.  
XX  
KW Probe; human; breast disease; breast cancer; development disorder; ss;  
KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.  
XX  
OS Homo sapiens.  
XX  
PN WO200157270-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 29-JAN-2001; 2001WO-US00661.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX



PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2001-476286/51.  
 XX  
 PT Novel single exon nucleic acid probe used to measuring gene expression  
 PT in a human breast -  
 XX  
 PS Claim 25; SEQ ID No 9904; 322pp; English.  
 XX  
 CC The present invention relates to novel single exon nucleic acid probes.  
 CC The present sequence is one such probe. The probes are useful for  
 CC measuring human gene expression in a human breast sample, where the probe  
 CC hybridises at high stringency to a nucleic acid expressed in the human  
 CC breast. The probes are useful for predicting, diagnosing, grading,  
 CC staging, monitoring and prognosing diseases of the human breast,  
 CC particularly those diseases with polygenic aetiology. The diseases  
 CC include: breast cancer, disorders of development, inflammatory diseases  
 CC of the breast, fibrocystic changes, proliferative breast disease and  
 CC non-carcinoma tumours.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 273 BP; 65 A; 54 C; 71 G; 83 T; 0 other;

Query Match 7.3%; Score 32.4; DB 22; Length 273;  
 Best Local Similarity 56.6%; Pred. No. 0.62;  
 Matches 60; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

Qy 71 tgccttccaagaatcctccaaccacccttggtgcccaagcaagccacaaaaagtgggcaa 130  
 ||||| ||| || || | ||| | || | ||  
 Db 173 TGCCTTCCCCACCTCCCCCTCTTCTCCCTCCTCTACTACCAGCCCCCATAAACAGACAG 114  
 Qy 131 accaaagaggacagcaatgctaggaaaatgacgatgacaaagacga 176  
 | | | ||| |||| |||| | || | || | ||  
 Db 113 AAGACAAAGGAGTTCAATGTGAGGAAGAGGAAGAAGAGAAGAAAGA 68

RESULT 14  
 AAQ49754/c  
 ID AAQ49754 standard; DNA; 7607 BP.  
 XX  
 AC AAQ49754;  
 XX  
 DT 10-MAR-1994 (first entry)  
 XX  
 DE pTK gene LpTK-2.  
 XX  
 KW pTK; protein tyrosine kinase; catalytic domain; c-kit; megakaryocyte;  
 KW lymphocyte; amplification; primer; polymerase chain reaction; PCR; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers

FT CDS 1858..3375  
 FT /\*tag= a  
 XX  
 PN WO9315201-A.  
 XX  
 PD 05-AUG-1993.  
 XX  
 PF 22-JAN-1993; 93WO-US00586.  
 XX  
 PR 22-JAN-1992; 92US-0826935.  
 XX  
 PA (NEWE-) NEW ENGLAND DEACONESS HOSPITAL.  
 XX  
 PI Avraham H, Cowley S, Groopman J, Scadden D;  
 XX  
 DR WPI; 1993-320330/40.  
 DR P-PSDB; AAR41941.  
 XX  
 PT New protein tyrosine kinase genes and proteins encoded by genes -  
 PT are of human mega-karyocytic origin  
 XX  
 PS Claim 2; Fig 5; 60pp; English.  
 XX  
 CC pTK genes were identified using two sets of degenerative  
 CC oligonucleotide primers: a first set which amplifies all pTK DNA  
 CC segments (AAQ49743-44), and a second set which amplifies highly  
 CC conserved sequences present in the catalytic domain of the c-kit  
 CC subgroup of pTKs (AAQ49745-46). The pTK genes identified are described  
 CC in AAQ49747-57 and AAR41897-02.  
 CC The LpTKs are expressed in lymphocytic cells, as well as  
 CC megakaryocytic cells. The partial and full-length LpTK2 gene  
 CC sequences are given in AAQ49749 and AAQ49754 respectively. The  
 CC protein sequence corresp. to AAQ49749 is claimed (claim 7) and  
 CC stated as given in the specification, however is missing from  
 CC the publication.  
 XX  
 SQ Sequence 7607 BP; 1953 A; 1851 C; 1694 G; 2109 T; 0 other;

Query Match 7.2%; Score 31.8; DB 14; Length 7607;  
 Best Local Similarity 50.3%; Pred. No. 4.5;  
 Matches 78; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

Qy 217 gccatgctatctagattccggtcttaatgagtacgtctgcagaaagactaataagtgcta 276  
 | | | | | | | | | | | | | | | | | | | | | |  
 Db 2207 GACACGAAATAAAGCTGCCGGTGAAGTGGACTGCGCCGAAGCCATTTCGTAGTAATAAAT 2148  
  
 Qy 277 taagagcttggtgctctgcgtggcgagttgtcaaccatcatcatgaattcaagatactgc 336  
 | | | | | | | | | | | | | | | | | | | | | |  
 Db 2147 TCAGCATTAAGTCCGATGTATGGTCATTTGGAATCCTTCTTTATGAAATCATTACTTATG 2088  
  
 Qy 337 ggagacatcatgatactgcggagacagacggccag 371  
 | | | | | | | | | | | | | | | | | | | | | |  
 Db 2087 GCAAATGCCTTACAGTGGTATGACAGGTGCCAG 2053

AAT03097/c  
ID AAT03097 standard; DNA; 7607 BP.  
XX  
AC AAT03097;  
XX  
DT 14-FEB-1996 (first entry)  
XX  
DE Protein tyrosine-kinase LpTK2 gene.  
XX  
KW Protein tyrosine-kinase; pTK; LpTK2; agonist; cell growth;  
KW differentiation; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO9527061-A1.  
XX  
PD 12-OCT-1995.  
XX  
PF 04-APR-1995; 95WO-US04228.  
XX  
PR 04-APR-1994; 94US-0222616.  
XX  
PA (GETH ) GENENTECH INC.  
XX  
PI Bennett BD, Goeddel D, Lee JM, Matthews W, Tsai SP;  
PI Wood WI;  
XX  
DR WPI; 1995-366160/47.  
DR P-PSDB; AAR85929.  
XX  
PT Agonist antibodies which activate specific protein tyrosine  
PT kinase(s) - also activate chimeric proteins of kinase extracellular  
PT domain and Ig constant domain, useful for studying, and therapeutic  
PT modulation of, cell growth and differentiation  
XX  
PS Disclosure; Page 48-56; 125pp; English.  
XX  
CC DNA probes based on protein tyrosine-kinase (pTK) sequences were used  
CC to screen cDNA libraries to identify novel pTK genes. A LpTK2 gene  
CC (AAT03097) was isolated from lymphocytic and megakaryocytic cell  
CC libraries. The gene can be used to produce recombinant LpTK2, to  
CC identify other new pTK genes, or to design drugs, peptides or  
CC antisense constructs that modulate pTK activity.  
XX  
SQ Sequence 7607 BP; 1954 A; 1851 C; 1693 G; 2109 T; 0 other;

Query Match 7.2%; Score 31.8; DB 16; Length 7607;  
Best Local Similarity 50.3%; Pred. No. 4.5;  
Matches 78; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

Qy 217 gccatgctatctagattccgggtcttaatgagtacgtctgcagaaagactaataagtgc 276  
| | | . | | | | | | | | | | | | | | | | | |  
Db 2207 GACACGAAATAAAGCTGCCGGTGAAGTGGACTGCGCCCGAAGCCATTTCGTAGTAATAAAT 2148  
Qy 277 taagagcttggtgctctgcgtggcgagttgtcaaccatcatcatgaattcaagatactgc 336  
| | | | | | | | | | | | | | | | | | | | |

Db 2147 TCAGCATTAAGTCCGATGTATGGTCATTTGGAATCCTTCTTTATGAAATCATTACTTATG 2088

Qy 337 ggagacatcatgatactgcgagacagacggccag 371

| | | | | | | | | | | | | | | |

Db 2087 GCAAATGCCTTACAGTGGTATGACAGGTGCCAG 2053

Search completed: February 7, 2002, 10:59:55

Job time: 4981 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 10:51:40 ; Search time 172.96 Seconds  
(without alignments)  
581.384 Million cell updates/sec

Title: US-09-394-745-6154  
Perfect score: 444  
Sequence: 1 cgaaaacactggtacccaaa.....tcccattttaagaaataaat 444

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result	No.	Score	Query Match	Length	DB	ID	Description
c	1	41.4	9.3	7218	1	US-08-232-463-14	Sequence 14, Appl

c	2	34.2	7.7	1404	1	US-08-204-656B-3	Sequence 3, Appli
c	3	34.2	7.7	1404	1	US-08-204-656B-5	Sequence 5, Appli
c	4	34.2	7.7	1404	1	US-08-204-656B-7	Sequence 7, Appli
c	5	34.2	7.7	1404	1	US-08-470-702-2	Sequence 2, Appli
c	6	34.2	7.7	1404	1	US-08-470-702-3	Sequence 3, Appli
c	7	34.2	7.7	1404	1	US-08-470-702-4	Sequence 4, Appli
c	8	34.2	7.7	1404	1	US-08-467-831-2	Sequence 2, Appli
c	9	34.2	7.7	1404	1	US-08-467-831-3	Sequence 3, Appli
c	10	34.2	7.7	1404	1	US-08-467-831-4	Sequence 4, Appli
c	11	32.6	7.3	1404	1	US-08-204-656B-1	Sequence 1, Appli
c	12	32.6	7.3	1404	1	US-08-470-702-1	Sequence 1, Appli
c	13	32.6	7.3	1404	1	US-08-467-831-1	Sequence 1, Appli
	14	31.8	7.2	2770	5	PCT-US95-05008-5	Sequence 5, Appli
c	15	31.8	7.2	7607	1	US-08-222-616-19	Sequence 19, Appl
c	16	31.8	7.2	7607	5	PCT-US95-04228-19	Sequence 19, Appl
	17	30.8	6.9	2277	1	US-08-676-967-2	Sequence 2, Appli
	18	30.8	6.9	2277	1	US-08-676-974-2	Sequence 2, Appli
	19	30.8	6.9	2277	2	US-09-098-487-2	Sequence 2, Appli
	20	30.2	6.8	289	4	US-09-007-005-17	Sequence 17, Appl
	21	30.2	6.8	289	4	US-09-244-796-17	Sequence 17, Appl
	22	29.8	6.7	2295	1	US-08-375-300-3	Sequence 3, Appli
	23	29.8	6.7	2295	3	US-09-177-431-3	Sequence 3, Appli
	24	29.8	6.7	2295	5	PCT-US95-16930-3	Sequence 3, Appli
	25	29.8	6.7	4080	1	US-08-375-300-1	Sequence 1, Appli
	26	29.8	6.7	4080	3	US-09-177-431-1	Sequence 1, Appli
	27	29.8	6.7	4080	5	PCT-US95-16930-1	Sequence 1, Appli
	28	29	6.5	1212	2	US-09-092-770-18	Sequence 18, Appl
	29	29	6.5	1212	4	US-09-222-851-18	Sequence 18, Appl
	30	28.8	6.5	246240	2	US-08-724-394A-20	Sequence 20, Appl
	31	28.8	6.5	246240	2	US-08-724-394A-21	Sequence 21, Appl
	32	28.8	6.5	246240	2	US-08-724-394A-22	Sequence 22, Appl
c	33	28.6	6.4	375	2	US-08-633-879C-5	Sequence 5, Appli
c	34	28.2	6.4	654	4	US-08-998-416-176	Sequence 176, App
c	35	28.2	6.4	4038	3	US-08-969-125-8	Sequence 8, Appli
	36	28.2	6.4	5438	4	US-08-456-200B-5	Sequence 5, Appli
c	37	28	6.3	771	4	US-09-277-716-19	Sequence 19, Appl
c	38	28	6.3	1113	3	US-09-043-627-9	Sequence 9, Appli
c	39	28	6.3	1146	4	US-09-277-716-21	Sequence 21, Appl
	40	28	6.3	4122	4	US-09-321-831-1	Sequence 1, Appli
	41	27.8	6.3	1207	1	US-08-362-670B-3	Sequence 3, Appli
	42	27.8	6.3	1207	3	US-08-333-576C-3	Sequence 3, Appli
	43	27.8	6.3	1207	3	US-08-289-222E-1	Sequence 1, Appli
	44	27.8	6.3	1207	4	US-09-054-526B-1	Sequence 1, Appli
	45	27.8	6.3	1207	4	US-08-808-324-3	Sequence 3, Appli

#### ALIGNMENTS

RESULT 1  
 US-08-232-463-14/c  
 ; Sequence 14, Application US/08232463  
 ; Patent No. 5670367  
 ; GENERAL INFORMATION:  
 ; APPLICANT: DORNER, F.  
 ; APPLICANT: SCHEIFLINGER, F.  
 ; APPLICANT: FALKNER, F. G.



[illegible]

RESULT 2

```

;     LENGTH: 1404 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: double
;     TOPOLOGY: linear
;     MOLECULE TYPE: other nucleic acid
;     DESCRIPTION: /desc = "Synthetic nucleic acid"
;     HYPOTHETICAL: NO
;     ANTI-SENSE: NO
;     IMMEDIATE SOURCE:
;     CLONE: Derived from plasmid pSf`1 (Agric. Biol. Chem.
;     FEATURE:
;     NAME/KEY: CDS
;     LOCATION: 1..1404
;     OTHER INFORMATION: /note= "Nucleotides 1-1404
;     OTHER INFORMATION: correspond to nucleotides 79-1482 in the
Saccharomycopsis
;     OTHER INFORMATION: fibuligera -amylase structural gene"
US-08-204-656B-3

```

```

Query Match          7.7%; Score 34.2; DB 1; Length 1404;
Best Local Similarity 49.7%; Pred. No. 0.046;
Matches 87; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

```

```

Qy      51 cctccaaagaataatccgggtgccttccaagaatcctccaaccacccttggtgcccaagc 110
      ||| | || | ||| | ||| | | |||| | | | ||||| | || |
Db      689 CCTACTGAGTAAACTCCAGATGCACTAACAAAATCCGGGAAAAAGCCTTGGTCCACATGT 630

Qy     111 aagccacaaaaagtgggcaaaccaaagaggacagcaatgctaggaaaatgacgatgacaa 170
      ||| | | | ||||| | | ||| | |||| | | ||
Db     629 TTAGCACTATCAATTCTTAAACCATCAATTGAGTAATTGCCAACAAAATCTTTAACCCAA 570

Qy     171 agacgagggcatcgggcaacatacttgtagccgtaatgacgacggggccatgcta 225
      | | | ||| || | | | |||| | | | ||| |||| |
Db     569 GAATTGAAAAC TGAGGCCACGTCGCTATCTTCCGTTCTCAAATCTGGTAATGCAA 515

```

```

RESULT 3
US-08-204-656B-5/c
; Sequence 5, Application US/08204656B
; Patent No. 5538882
; GENERAL INFORMATION:
; APPLICANT: Matsui, Ikuo
; APPLICANT: Ishikawa, Kazuhiko
; APPLICANT: Miyairi, Sachio
; APPLICANT: Honda, Koichi
; TITLE OF INVENTION: Variant-Type Carbohydrate Hydrolase,
; TITLE OF INVENTION: Variant Gene Of The Enzyme And Method For Producing
; TITLE OF INVENTION: Oligosaccharide Using The Enzyme
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP
; STREET: 8110 Gatehouse Road, Suite 500 East
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22042

```



```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/204,656B
; FILING DATE: 02-MAR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiner, Marc S.
; REGISTRATION NUMBER: 32,181
; REFERENCE/DOCKET NUMBER: 234-252P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1404 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "Synthetic DNA"
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; IMMEDIATE SOURCE:
; CLONE: Derived from plasmid pSf`1 (Agric. Biol. Chem.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1404
; OTHER INFORMATION: /note= "Nucleotides 1-1404
; OTHER INFORMATION: correspond to nucleotides 79-1482 of the
Saccharomycopsis
; OTHER INFORMATION: fibuligera -amylase structural gene"
US-08-204-656B-5

```

```

Query Match          7.7%; Score 34.2; DB 1; Length 1404;
Best Local Similarity 49.7%; Pred. No. 0.046;
Matches 87; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

```

```

Qy      51 cctccaaagaataatccgggtgccttccaagaatcctccaaccacccttggtgcccaagc 110
      ||| | || | ||| | ||| | | |||| | | | ||||| | || |
Db      689 CCTACTGAGTAAACTCCAGATGCACTAACAAAATCCGGGAAAAAGCCTTGGTCCACATGT 630

Qy     111 aagccacaaaaagtgggcaaaccaaagaggacagcaatgctaggaaaatgacgatgacaa 170
      ||| | | | ||||| | | ||| | |||| | | |||
Db     629 TTAGCACTATCAATTCTTAAACCATCAATTGAGTAATTGCCAACAAAATCTTTAACCCAA 570

Qy     171 agacgagggcatcgggcaacatacttgtagccgtaatgacgacgggcatgcta 225
      | | | ||| || | | | |||| | | | ||| ||| |
Db     569 GAATTGAAACTGAGGCCACGTCGCTATCTTCCGTTCTCAAATCTGGTAATGCAA 515

```

RESULT 4

US-08-204-656B-7/c  
 ; Sequence 7, Application US/08204656B  
 ; Patent No. 5538882  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Matsui, Ikuo  
 ; APPLICANT: Ishikawa, Kazuhiko  
 ; APPLICANT: Miyairi, Sachio  
 ; APPLICANT: Honda, Koichi  
 ; TITLE OF INVENTION: Variant-Type Carbohydrate Hydrolase,  
 ; TITLE OF INVENTION: Variant Gene Of The Enzyme And Method For Producing  
 ; TITLE OF INVENTION: Oligosaccharide Using The Enzyme  
 ; NUMBER OF SEQUENCES: 18  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP  
 ; STREET: 8110 Gatehouse Road, Suite 500 East  
 ; CITY: Falls Church  
 ; STATE: Virginia  
 ; COUNTRY: U.S.A.  
 ; ZIP: 22042  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/204,656B  
 ; FILING DATE: 02-MAR-1994  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Weiner, Marc S.  
 ; REGISTRATION NUMBER: 32,181  
 ; REFERENCE/DOCKET NUMBER: 234-252P  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (703) 205-8000  
 ; TELEFAX: (703) 205-8050  
 ; TELEX: 248345  
 ; INFORMATION FOR SEQ ID NO: 7:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1404 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: other nucleic acid  
 ; DESCRIPTION: /desc = "Synthetic DNA"  
 ; HYPOTHETICAL: NO  
 ; ANTI-SENSE: NO  
 ; IMMEDIATE SOURCE:  
 ; CLONE: Derived from plasmid pSf`1 (Agric. Biol. Chem.  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 1..1404  
 ; OTHER INFORMATION: /note= "Nucleotides 1-1404  
 ; OTHER INFORMATION: correspond to nucleotides 79-1482 of the  
 Saccharomycopsis  
 ; OTHER INFORMATION: fibuligera -amylase structural gene"  
 US-08-204-656B-7

Query Match 7.7%; Score 34.2; DB 1; Length 1404;  
Best Local Similarity 49.7%; Pred. No. 0.046;  
Matches 87; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

```
Qy 51 cctccaaagaataatccgggtgccttccaagaatcctccaaccacccttggtgccaagc 110
    ||| | || | || | || | || | || | || | || | || | || | || |
Db 689 CCTACTGAGTAAACTCCAGATGCTAATAACAAATCCGGGAAAAGCCTTGGTCCACATGT 630

Qy 111 aagccacaaaaagtgggcaaaccaaagaggacagcaatgctaggaaaatgacgatgacaa 170
    ||| | | | |||| | | || | || | || | || | || |
Db 629 TTAGCACTATCAATTCTTAAACCATCAATTGAGTAATTGCCAACAAATCTTTAACCCAA 570

Qy 171 agacgagggcatcgggcaacatacttgtagccgtaatgacgacgggcatgcta 225
    | | | | | | | | | | | | | | | | | | | | | |
Db 569 GAATTGAAAACCTGAGGCCACGTCGCTATCTCCGTTCTCAAATCTGGTAATGCAA 515
```

RESULT 5  
US-08-470-702-2/c  
; Sequence 2, Application US/08470702  
; Patent No. 5631149  
; GENERAL INFORMATION:  
; APPLICANT: MATSUI, IKUO  
; APPLICANT: ISHIKAWA, KAZUHIKO  
; APPLICANT: MIYAIRI, SACHIO  
; APPLICANT: HONDA, KOICHI  
; TITLE OF INVENTION: VARIANT-TYPE CARBOHYDRATE HYDROLASE,  
; TITLE OF INVENTION: VARIANT GENE OF THE ENZYME AND METHOD FOR PRODUCING  
; TITLE OF INVENTION: OLIGOSACCHARIDE USING THE ENZYME  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH  
; STREET: 8110 Gatehouse Road, Suite 500 East  
; CITY: Falls Church  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22042  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/470,702  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/204,656  
; FILING DATE: 02-MAR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: WEINER, MARC S.  
; REGISTRATION NUMBER: 32,181  
; REFERENCE/DOCKET NUMBER: 234-252P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 205-8000  
; TELEFAX: (703) 205-8050

```

;       TELEX: 248345
; INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 1404 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: double
;     TOPOLOGY: linear
;   MOLECULE TYPE: DNA (synthetic)
;   HYPOTHETICAL: NO
;   ANTI-SENSE: NO
US-08-470-702-2

```

```

Query Match          7.7%; Score 34.2; DB 1; Length 1404;
Best Local Similarity 49.7%; Pred. No. 0.046;
Matches 87; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

```

```

Qy      51 cctccaaagaataatccgggtgccttccaagaatcctccaaccacccttggtgcccaagc 110
      ||| | || | | || | ||| | | |||| | | | ||||| | || |
Db      689 CCTACTGAGTAAACTCCAGATGCACTAACAAAATCCGGGAAAAAGCCTTGGTCCACATGT 630

Qy     111 aagccacaaaaagtgggcaaaccaaagaggacagcaatgctaggaaaatgacgatgacaa 170
      ||| | | | | ||||| | | ||| | |||| | | |||
Db     629 TTAGCACTATCAATTCTTAAACCATCAATTGAGTAATTGCCAACAAAATCTTTAACCCAA 570

Qy     171 agacgagggcatcgggcaacatacttggttagccgtaatgacgacggggccatgcta 225
      | | | ||| | | | ||| | | | ||| ||| |
Db     569 GAATTGAAAACAGGCGCAGCTCGCTATCTTCCGTTCTCAAATCTGGTAATGCAA 515

```

```

RESULT 6
US-08-470-702-3/c
; Sequence 3, Application US/08470702
; Patent No. 5631149
; GENERAL INFORMATION:
;   APPLICANT: MATSUI, IKUO
;   APPLICANT: ISHIKAWA, KAZUHIKO
;   APPLICANT: MIYAIRI, SACHIO
;   APPLICANT: HONDA, KOICHI
;   TITLE OF INVENTION: VARIANT-TYPE CARBOHYDRATE HYDROLASE,
;   TITLE OF INVENTION: VARIANT GENE OF THE ENZYME AND METHOD FOR PRODUCING
;   TITLE OF INVENTION: OLIGOSACCHARIDE USING THE ENZYME
;   NUMBER OF SEQUENCES: 17
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
;     STREET: 8110 Gatehouse Road, Suite 500 East
;     CITY: Falls Church
;     STATE: Virginia
;     COUNTRY: U.S.A.
;     ZIP: 22042
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/470,702

```

```

; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/204,656
; FILING DATE: 02-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: WEINER, MARC S.
; REGISTRATION NUMBER: 32,181
; REFERENCE/DOCKET NUMBER: 234-252P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1404 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (synthetic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-470-702-3

```

```

Query Match          7.7%; Score 34.2; DB 1; Length 1404;
Best Local Similarity 49.7%; Pred. No. 0.046;
Matches 87; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

```

```

Qy      51 cctccaaagaataatccgggtgccttccaagaatcctccaaccacccttggtgcccaagc 110
      ||| | || | || | || | || | || | || | || | || | || |
Db      689 CCTACTGAGTAAACTCCAGATGCACTAACAAAATCCGGGAAAAGCCTTGGTCCACATGT 630

Qy     111 aagccacaaaaagtgggcaaaccaaagaggacagcaatgctaggaaaatgacgatgacaa 170
      ||| | | | || | || | || | || | || | || | || | || |
Db     629 TTAGCACTATCAATTCTTAAACCATCAATTGAGTAATTGCCAACAAAATCTTTAACCCAA 570

Qy     171 agacgagggcatcgggcaacataacttgtagccgtaatgacgacggggccatgcta 225
      | | | || | || | || | || | || | || | || | || |
Db     569 GAATTGAAAACCTGAGGCCACGTCGCTATCTTCCGTTCTCAAATCTGGTAATGCAA 515

```

```

RESULT 7
US-08-470-702-4/c
; Sequence 4, Application US/08470702
; Patent No. 5631149
; GENERAL INFORMATION:
; APPLICANT: MATSUI, IKUO
; APPLICANT: ISHIKAWA, KAZUHIKO
; APPLICANT: MIYAIRI, SACHIO
; APPLICANT: HONDA, KOICHI
; TITLE OF INVENTION: VARIANT-TYPE CARBOHYDRATE HYDROLASE,
; TITLE OF INVENTION: VARIANT GENE OF THE ENZYME AND METHOD FOR PRODUCING
; TITLE OF INVENTION: OLIGOSACCHARIDE USING THE ENZYME
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH

```

```

; STREET: 8110 Gatehouse Road, Suite 500 East
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22042
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,702
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/204,656
; FILING DATE: 02-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: WEINER, MARC S.
; REGISTRATION NUMBER: 32,181
; REFERENCE/DOCKET NUMBER: 234-252P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1404 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (synthetic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-470-702-4

```

```

Query Match          7.7%; Score 34.2; DB 1; Length 1404;
Best Local Similarity 49.7%; Pred. No. 0.046;
Matches 87; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

```

```

Qy 51 cctccaaagaataatccgggtgccttccaagaatcctccaaccacccttggtgcccaagc 110
    ||| | || | ||| | ||| | | |||| | | | ||||| | || |
Db 689 CCTACTGAGTAAACTCCAGATGCACTAACAAAATCCGGGAAAAAGCCTTGGTCCACATGT 630

Qy 111 aagccacaaaaagtgggcaaaccaaagaggacagcaatgctaggaaaatgacgatgacaa 170
    ||| | | | ||||| | | ||| | |||| | | |||
Db 629 TTAGCACTATCAATTCTTAAACCATCAATTGAGTAATTGCCAACAAAATCTTTAACCCAA 570

Qy 171 agacgagggcatcgggcaacatacttggttagccgtaatgacgacgggccaatgcta 225
    | | | | ||| || | | | |||| | | | ||| |||| |
Db 569 GAATTGAAACTGAGGCCACGTCGCTATCTTCCGTTCTCAAATCTGGTAATGCAA 515

```

```

RESULT 8
US-08-467-831-2/c
; Sequence 2, Application US/08467831

```

```

; Patent No. 5635378
; GENERAL INFORMATION:
;   APPLICANT: MATSUI, IKUO
;   APPLICANT: ISHIKAWA, KAZUHIKO
;   APPLICANT: MIYAIRI, SACHIO
;   APPLICANT: HONDA, KOICHI
;   TITLE OF INVENTION: VARIANT-TYPE CARBOHYDRATE HYDROLASE,
;   TITLE OF INVENTION: VARIANT GENE OF THE ENZYME AND METHOD FOR PRODUCING
;   TITLE OF INVENTION: OLIGOSACCHARIDE USING THE ENZYME
;   NUMBER OF SEQUENCES: 17
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
;     STREET: 8110 Gatehouse Road, Suite 500 East
;     CITY: Falls Church
;     STATE: Virginia
;     COUNTRY: U.S.A.
;     ZIP: 22042
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/467,831
;     FILING DATE: 06-JUN-1995
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/204,656
;     FILING DATE: 02-MAR-1994
;   ATTORNEY/AGENT INFORMATION:
;     NAME: WEINER, MARC S.
;     REGISTRATION NUMBER: 32,181
;     REFERENCE/DOCKET NUMBER: 234-252P
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (703) 205-8000
;     TELEFAX: (703) 205-8050
;     TELEX: 248345
;   INFORMATION FOR SEQ ID NO: 2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH: 1404 base pairs
;       TYPE: nucleic acid
;       STRANDEDNESS: double
;       TOPOLOGY: linear
;     MOLECULE TYPE: DNA (synthetic)
;     HYPOTHETICAL: NO
;     ANTI-SENSE: NO
US-08-467-831-2

```

```

Query Match          7.7%; Score 34.2; DB 1; Length 1404;
Best Local Similarity 49.7%; Pred. No. 0.046;
Matches 87; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

```

```

Qy      51 cctccaaagaataatccgggtgccttccaagaatcctccaaccacccttggtgcccaagc 110
      ||| | || | ||| | ||| | | |||| | | | ||||| | || |
Db      689 CCTACTGAGTAAACTCCAGATGCACTAACAAAATCCGGGAAAAAGCCTTGGTCCACATGT 630

```

Qy 111 aagccacaaaaagtgggcaaaccaaagaggacagcaatgctaggaaaatgacgatgacaa 170  
 ||| | | | |||| | | ||| | |||| | |||  
 Db 629 TTAGCACTATCAATTCTTAAACCATCAATTGAGTAATTGCCAACAAAATCTTTAACCCAA 570

Qy 171 agacgagggcatcgggcaacatacttgtagccgtaatgacgacggggccatgcta 225  
 | | ||| || | | |||| | | | ||| |||| |  
 Db 569 GAATTGAAAACCTGAGGCCACGTCGCTATCTTCCGTTCTCAAATCTGGTAATGCAA 515

RESULT 9  
 US-08-467-831-3/c  
 ; Sequence 3, Application US/08467831  
 ; Patent No. 5635378  
 ; GENERAL INFORMATION:  
 ; APPLICANT: MATSUI, IKUO  
 ; APPLICANT: ISHIKAWA, KAZUHIKO  
 ; APPLICANT: MIYAIRI, SACHIO  
 ; APPLICANT: HONDA, KOICHI  
 ; TITLE OF INVENTION: VARIANT-TYPE CARBOHYDRATE HYDROLASE,  
 ; TITLE OF INVENTION: VARIANT GENE OF THE ENZYME AND METHOD FOR PRODUCING  
 ; TITLE OF INVENTION: OLIGOSACCHARIDE USING THE ENZYME  
 ; NUMBER OF SEQUENCES: 17  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH  
 ; STREET: 8110 Gatehouse Road, Suite 500 East  
 ; CITY: Falls Church  
 ; STATE: Virginia  
 ; COUNTRY: U.S.A.  
 ; ZIP: 22042  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/467,831  
 ; FILING DATE: 06-JUN-1995  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/204,656  
 ; FILING DATE: 02-MAR-1994  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: WEINER, MARC S.  
 ; REGISTRATION NUMBER: 32,181  
 ; REFERENCE/DOCKET NUMBER: 234-252P  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (703) 205-8000  
 ; TELEFAX: (703) 205-8050  
 ; TELEX: 248345  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1404 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (synthetic)  
 ; HYPOTHETICAL: NO



; ANTI-SENSE: NO  
US-08-467-831-3

Query Match 7.7%; Score 34.2; DB 1; Length 1404;  
Best Local Similarity 49.7%; Pred. No. 0.046;  
Matches 87; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

```
Qy      51 cctccaaagaataatccgggtgccttccaagaatcctccaaccacccttggtgcccaagc 110
          ||| | || | || | || | || | || | || | || | || | || | || |
Db      689 CCTACTGAGTAAACTCCAGATGCACTAACAAAATCCGGGAAAAAGCCTTGGTCCACATGT 630

Qy     111 aagccacaaaaagtgggcaaaccaaagaggacagcaatgctaggaaaatgacgatgacaa 170
          ||| | | | | || | || | || | || | || | || | || | || |
Db     629 TTAGCACTATCAATTCTTAAACCATCAATTGAGTAATTGCCAACAAAATCTTTAACCCAA 570

Qy     171 agacgagggcatcgggcaacatacttgtagccgtaatgacgacggggccatgcta 225
          | | | | | | | || | || | || | || | || | || | || | || |
Db     569 GAATTGAAAACCTGAGGCCACGTCGCTATCTTCCGTTCTCAAATCTGGTAATGCAA 515
```

RESULT 10

US-08-467-831-4/c

; Sequence 4, Application US/08467831

; Patent No. 5635378

; GENERAL INFORMATION:

; APPLICANT: MATSUI, IKUO

; APPLICANT: ISHIKAWA, KAZUHIKO

; APPLICANT: MIYAIRI, SACHIO

; APPLICANT: HONDA, KOICHI

; TITLE OF INVENTION: VARIANT-TYPE CARBOHYDRATE HYDROLASE,

; TITLE OF INVENTION: VARIANT GENE OF THE ENZYME AND METHOD FOR PRODUCING

; TITLE OF INVENTION: OLIGOSACCHARIDE USING THE ENZYME

; NUMBER OF SEQUENCES: 17

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH

; STREET: 8110 Gatehouse Road, Suite 500 East

; CITY: Falls Church

; STATE: Virginia

; COUNTRY: U.S.A.

; ZIP: 22042

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/467,831

; FILING DATE: 06-JUN-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/204,656

; FILING DATE: 02-MAR-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: WEINER, MARC S.

; REGISTRATION NUMBER: 32,181

; REFERENCE/DOCKET NUMBER: 234-252P

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1404 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (synthetic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-467-831-4

```

```

Query Match          7.7%; Score 34.2; DB 1; Length 1404;
Best Local Similarity 49.7%; Pred. No. 0.046;
Matches 87; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

```

```

Qy      51 cctccaaagaataatccgggtgccttccaagaatcctccaaccacccttggtgcccaagc 110
      ||| | || | ||| | ||| | | |||| | | | ||||| | || |
Db      689 CCTACTGAGTAAACTCCAGATGCACTAACAAAATCCGGGAAAAAGCCTTGGTCCACATGT 630

Qy     111 aagccacaaaaagtgggcaaaccaaagaggacagcaatgctaggaaaatgacgatgacaa 170
      ||| | | | ||||| | | ||| | |||| | | |||
Db      629 TTAGCACTATCAATTCTTAAACCATCAATTGAGTAATTGCCAACAAAATCTTTAACCCAA 570

Qy     171 agacgagggcatcgggcaacatacttggttagccgtaatgacgacgggcatgcta 225
      | | | ||| | | | ||| | | | ||| ||| |
Db      569 GAATTGAAAACCTGAGGCCACGTCGCTATCTTCCGTTCTCAAATCTGGTAATGCAA 515

```

# RESULT 11

US-08-204-656B-1/c

; Sequence 1, Application US/08204656B

; Patent No. 5538882

## ; GENERAL INFORMATION:

; APPLICANT: Matsui, Ikuo

; APPLICANT: Ishikawa, Kazuhiko

; APPLICANT: Miyairi, Sachio

; APPLICANT: Honda, Koichi

; TITLE OF INVENTION: Variant-Type Carbohydrate Hydrolase,

; TITLE OF INVENTION: Variant Gene Of The Enzyme And Method For Producing

; TITLE OF INVENTION: Oligosaccharide Using The Enzyme

; NUMBER OF SEQUENCES: 18

## ; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP

; STREET: 8110 Gatehouse Road, Suite 500 East

; CITY: Falls Church

; STATE: Virginia

; COUNTRY: U.S.A.

; ZIP: 22042

## ; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

```

;   SOFTWARE:  PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER:  US/08/204,656B
;   FILING DATE:  02-MAR-1994
;   CLASSIFICATION:  435
;   ATTORNEY/AGENT INFORMATION:
;   NAME:  Weiner, Marc S.
;   REGISTRATION NUMBER:  32,181
;   REFERENCE/DOCKET NUMBER:  234-252P
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE:  (703) 205-8000
;   TELEFAX:  (703) 205-8050
;   TELEX:  248345
;   INFORMATION FOR SEQ ID NO:  1:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH:  1404 base pairs
;   TYPE:  nucleic acid
;   STRANDEDNESS:  double
;   TOPOLOGY:  linear
;   MOLECULE TYPE:  other nucleic acid
;   DESCRIPTION:  /desc = "Synthetic DNA"
;   HYPOTHETICAL:  NO
;   ANTI-SENSE:  NO
;   IMMEDIATE SOURCE:
;   CLONE:  Derived from plasmid pSf`1 (Agric. Biol. Chem.
;   FEATURE:
;   NAME/KEY:  CDS
;   LOCATION:  1..1404
;   OTHER INFORMATION:  /note= "Nucleotides 1-1404
;   OTHER INFORMATION:  correspond to nucleotides 79-1482 in the
Saccharomycopsis
;   OTHER INFORMATION:  fibuligera `'-amylase structural gene"
US-08-204-656B-1

```

```

Query Match          7.3%;  Score 32.6;  DB 1;  Length 1404;
Best Local Similarity 49.1%;  Pred. No. 0.17;
Matches 86;  Conservative 0;  Mismatches 89;  Indels 0;  Gaps 0;

```

```

Qy      51 cctccaaagaataatccgggtgccttccaagaatcctccaaccacccttggtgcccaagc 110
      ||| | || | || | || | || | || | || | || | || | || | || | || |
Db      689 CCTACTGAGTAAACTCCAGATGGACTAACAAAATCCGGGAAAAAGCCTTGGTCCACATGT 630

Qy     111 aagccacaaaaagtgggcaaaccaagaggacagcaatgctaggaaaatgacgatgacaa 170
      ||| | | | | |||| | | | || | || | || | || | || | || |
Db     629 TTAGCACTATCAATTCTTAAACCATCAATTGAGTAATTGCCAACAAAATCTTTAACCCAA 570

Qy     171 agacgagggcatcgggcaacatacttgtagccgtaatgacgacggggccatgcta 225
      | | | || | | || | || | || | || | || | || | || | || |
Db     569 GAATTGAAAACCTGAGGCCACGTCGCTATCTTCCGTTCTCAAATCTGGTAATGCAA 515

```

```

RESULT 12
US-08-470-702-1/c
; Sequence 1, Application US/08470702
; Patent No. 5631149
; GENERAL INFORMATION:

```

```

; APPLICANT: MATSUI, IKUO
; APPLICANT: ISHIKAWA, KAZUHIKO
; APPLICANT: MIYAIRI, SACHIO
; APPLICANT: HONDA, KOICHI
; TITLE OF INVENTION: VARIANT-TYPE CARBOHYDRATE HYDROLASE,
; TITLE OF INVENTION: VARIANT GENE OF THE ENZYME AND METHOD FOR PRODUCING
; TITLE OF INVENTION: OLIGOSACCHARIDE USING THE ENZYME
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
; STREET: 8110 Gatehouse Road, Suite 500 East
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22042
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,702
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/204,656
; FILING DATE: 02-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: WEINER, MARC S.
; REGISTRATION NUMBER: 32,181
; REFERENCE/DOCKET NUMBER: 234-252P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1404 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (synthetic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-470-702-1

```

```

Query Match          7.3%; Score 32.6; DB 1; Length 1404;
Best Local Similarity 49.1%; Pred. No. 0.17;
Matches 86; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

```

```

Qy      51 cctccaaagaataatccgggtgccttccaagaatcctccaaccacccttggtgcccaagc 110
      ||| | ||| | ||| | || | |||| | | | ||||| | || |
Db      689 CCTACTGAGTAAACTCCAGATGGACTAACAAAATCCGGGAAAAAGCCTTGGTCCACATGT 630

Qy      111 aagccacaaaaagtgggcaaaccaaagaggacagcaatgctaggaaaatgacgatgacaa 170
      ||| | | | ||||| | | ||| | |||| | | |||

```



Query Match 7.3%; Score 32.6; DB 1; Length 1404;  
Best Local Similarity 49.1%; Pred. No. 0.17;  
Matches 86; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

```
Qy      51 cctccaaagaataatccgggtgccttccaagaatcctccaaccacccttggtgccaagc 110
          ||| | || | || | || | | | ||| | | | ||| || | |
Db      689 CCTACTGAGTAAACTCCAGATGGACTAACAAAATCCGGGAAAAAGCCTTGGTCCACATGT 630

Qy     111 aagccacaaaaagtgggcaaaccaaagaggacagcaatgctaggaaaatgacgatgacaa 170
          ||| | | | ||| || | | ||| | ||| | | | |||
Db     629 TTAGCACTATCAATTCTTAAACCATCAATTGAGTAATTGCCAACAAAATCTTTAACCCAA 570

Qy     171 agacgagggcatcgggcaacatacttggttagccgtaatgacgacgggcatgcta 225
          | | | || | | | | ||| | | | ||| || | |
Db     569 GAATTGAAACTGAGGCCACGTCGCTATCTTCCGTTCTCAAATCTGGTAATGCAA 515
```

RESULT 14

PCT-US95-05008-5

; Sequence 5, Application PC/TUS9505008

; GENERAL INFORMATION:

; APPLICANT: Sugen, Inc.

; APPLICANT: 515 Galveston Drive

; APPLICANT: Redwood City, California 94063-4720

; APPLICANT: United States of America

; APPLICANT: Wissenschaften E.V.

; APPLICANT: Hofgarten Str. 2

; APPLICANT: Munchen 80539

; APPLICANT: Germany

; TITLE OF INVENTION: Novel Megakaryocytic Protein Tyrosine

; TITLE OF INVENTION: Kinases

; NUMBER OF SEQUENCES: 21

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pennie & Edmonds

; STREET: 1155 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10036

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US95/05008

; FILING DATE: 24-APR-1995

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/232,545

; FILING DATE: 22-APR-1994

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Coruzzi, Laura A.

; REGISTRATION NUMBER: 30,742

```

; REFERENCE/DOCKET NUMBER: 7683-074
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)790-9090
; TELEFAX: (212)869-9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2770 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
PCT-US95-05008-5

```

```

Query Match          7.2%; Score 31.8; DB 5; Length 2770;
Best Local Similarity 50.3%; Pred. No. 0.45;
Matches 78; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

```

```

Qy 217 gccatgctatctagattccggtcttaatgagtacgtctgcagaaagactaataagtgcta 276
    | | | | | | | | | | | | | | | | | | | | | |
Db 1534 GACACGAAATAAAGCTGCCGGTGAAGTGGACTGCGCCCGAAGCCATTCGTAGTAATAAAT 1593

Qy 277 taagagcttggtgctctgcgtggcgagttgtcaaccatcatcatgaattcaagatactgc 336
    | | | | | | | | | | | | | | | | | | | | | |
Db 1594 TCAGCATTAAGTCCGATGTATGGTCATTTGGAATCCTTCTTTATGAAATCATTACTTATG 1653

Qy 337 ggagacatcatgatactgcggagacagacggccag 371
    | | | | | | | | | | | | | | | | | |
Db 1654 GCAAATGCCTTACAGTGGTATGACAGGTGCCCAG 1688

```

RESULT 15

US-08-222-616-19/c

; Sequence 19, Application US/08222616

; Patent No. 5635177

; GENERAL INFORMATION:

; APPLICANT: Bennett, Brian D.

; APPLICANT: Goeddel, David

; APPLICANT: Lee, James M.

; APPLICANT: Matthews, William

; APPLICANT: Tsai, Siao Ping

; APPLICANT: Wood, William I.

; TITLE OF INVENTION: PROTEIN TYROSINE KINASE AGONIST

; TITLE OF INVENTION: ANTIBODIES

; NUMBER OF SEQUENCES: 42

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Genentech, Inc.

; STREET: 460 Point San Bruno Blvd

; CITY: South San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94080

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

```

;   SOFTWARE:  patin (Genentech)
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER:  US/08/222,616
;   FILING DATE:  4-APR-1994
;   CLASSIFICATION:  530
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:  PCT/US93/00586
;   FILING DATE:  22-JAN-1993
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:  07/826935
;   FILING DATE:  22-JAN-1992
;   ATTORNEY/AGENT INFORMATION:
;   NAME:  Lee, Wendy M.
;   REGISTRATION NUMBER:
;   REFERENCE/DOCKET NUMBER:  821P2
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE:  415/225-1994
;   TELEFAX:  415/952-9881
;   TELEX:  910/371-7168
;   INFORMATION FOR SEQ ID NO:  19:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH:  7607 bases
;   TYPE:  nucleic acid
;   STRANDEDNESS:  single
;   TOPOLOGY:  linear
US-08-222-616-19

```

```

Query Match          7.2%;  Score 31.8;  DB 1;  Length 7607;
Best Local Similarity  50.3%;  Pred. No. 0.76;
Matches  78;  Conservative  0;  Mismatches  77;  Indels  0;  Gaps  0;

```

```

Qy   217 gccatgctatctagattccggtcttaatgagtacgtctgcagaaagactaataagtgcta 276
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   2207 GACACGAAATAAAGCTGCCGGTGAAGTGGACTGCGCCCGAAGCCATTTCGTAGTAATAAAT 2148

Qy   277 taagagcttggtgctctgcgtggcgagttgtcaaccatcatcatgaattcaagatactgc 336
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   2147 TCAGCATTAAGTCCGATGTATGGTCATTTGGAATCCTTCTTTATGAAATCATTACTTATG 2088

Qy   337 ggagacatcatgatactgcggagacagacggccag 371
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   2087 GCAAATGCCTTACAGTGGTATGACAGGTGCCAG 2053

```

```

Search completed: February  7, 2002, 10:51:46
Job time: 6072 sec

```

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

```

Run on:          February  7, 2002, 08:20:41 ; Search time 4942.22 Seconds
                (without alignments)

```



965.381 Million cell updates/sec

Title: US-09-394-745-6154  
Perfect score: 444  
Sequence: 1 cgaaaacactggtacccaaa.....tcccattttaagaaataaat 444

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*  
1: em\_estfun:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estom:\*  
5: em\_estpl:\*  
6: em\_estba:\*  
7: em\_estro:\*  
8: em\_estov:\*  
9: em\_htc:\*  
10: gb\_est1:\*  
11: gb\_est2:\*  
12: gb\_htc:\*  
13: gb\_gss:\*  
14: em\_gss\_fun:\*  
15: em\_gss\_hum:\*  
16: em\_gss\_inv:\*  
17: em\_gss\_pln:\*  
18: em\_gss\_pro:\*  
19: em\_gss\_rod:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

		%		Query			Description
Result	No.	Score	Match	Length	DB	ID	
c	1	156.4	35.2	261	11	BG837106	BG837106 Zm08_06h0
c	2	121.2	27.3	129	10	AI673919	AI673919 605039B05
	3	45	10.1	309	11	BG240000	BG240000 OV1_31_G0
	4	45	10.1	435	11	BG240586	BG240586 OV1_31_G0
	5	44.6	10.0	1068	11	BG326023	BG326023 602424785

c	6	41	9.2	1159	13	CNS015XR	AL106041	Drosophil
	7	40.6	9.1	674	10	AW957338	AW957338	EST369528
c	8	39.4	8.9	1101	13	CNS00L6W	AL068145	Drosophil
	9	38.6	8.7	771	13	AQ210844	AQ210844	HS_2230_A
	10	38.2	8.6	1642	11	BG034748	BG034748	602301702
	11	38	8.6	697	11	BF671369	BF671369	602151249
	12	37.8	8.5	937	13	CNS006ST	AL065880	Drosophil
c	13	37.6	8.5	922	13	CNS0073W	AL066784	Drosophil
c	14	37.6	8.5	1128	13	CNS04GFJ	AL289576	Tetraodon
	15	37.2	8.4	861	13	CNS0075A	AL066834	Drosophil
c	16	37	8.3	947	11	BG614832	BG614832	602642367
c	17	36.8	8.3	750	13	AZ133062	AZ133062	OSJNBb010
	18	36.8	8.3	1025	13	CNS015OW	AL105722	Drosophil
c	19	36.8	8.3	1101	13	CNS00FXE	AL071370	Drosophil
	20	36.6	8.2	515	11	BF340131	BF340131	602037293
c	21	36.6	8.2	902	13	CNS006QP	AL065804	Drosophil
	22	36.6	8.2	1090	11	BF168976	BF168976	601775246
c	23	36.4	8.2	506	10	AW386337	AW386337	RC5-PT000
	24	36.4	8.2	939	13	CNS00CNG	AL059400	Drosophil
	25	36	8.1	587	10	AL515343	AL515343	AL515343
	26	36	8.1	979	11	BI116169	BI116169	602866385
	27	36	8.1	1226	11	BG114348	BG114348	602285924
	28	36	8.1	1613	11	BF341540	BF341540	602013546
	29	35.8	8.1	228	10	AI986114	AI986114	wr80e03.x
c	30	35.8	8.1	434	11	N45814	N45814	T1547 MVAT4
	31	35.8	8.1	669	10	BE661391	BE661391	0-E8 Gmax
	32	35.8	8.1	861	10	BE661431	BE661431	1105 Gmax
	33	35.8	8.1	1088	10	BE785449	BE785449	601478008
	34	35.2	7.9	419	10	AI748066	AI748066	sb47g09.y
	35	35.2	7.9	449	11	BG650798	BG650798	sad91g01.
	36	35.2	7.9	482	10	AW397798	AW397798	sg68e04.y
	37	35.2	7.9	485	10	AW395539	AW395539	sg72d07.y
	38	35.2	7.9	487	10	AW318042	AW318042	sg60d10.y
	39	35.2	7.9	488	11	BF595158	BF595158	su75g06.y
	40	35.2	7.9	491	10	AW397523	AW397523	sg79h04.y
	41	35.2	7.9	498	11	BF596382	BF596382	su71e11.y
	42	35.2	7.9	508	10	BE661447	BE661447	335 GmaxS
	43	35.2	7.9	513	10	AI941183	AI941183	sb85f09.y
	44	35.2	7.9	531	10	AI735787	AI735787	sb19a03.y
c	45	35.2	7.9	543	10	AW471784	AW471784	si15g12.y

# ALIGNMENTS

RESULT 1

BG837106/c

LOCUS BG837106 261 bp mRNA EST 25-MAY-2001

DEFINITION Zm08\_06h09\_A

Zm08\_AAFC\_ECORC\_Fusarium\_graminearum\_inoculated\_corn\_ear Zea mays

cDNA clone Zm08\_06h09, mRNA sequence.

ACCESSION BG837106

VERSION BG837106.1 GI:14203429

KEYWORDS EST.

SOURCE Zea mays.

ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.
REFERENCE	1 (bases 1 to 261)
AUTHORS	Harris,L.J., Balcerzak,M., Allard,S., Saparno,A., Couroux,P., De Moors,A., Hattori,J.I., Ouellet,T., Robert,L.S., Singh,J.A, Sprott ,D. and Tinker,N.A.
TITLE	Expressed Sequence Tags from Developing Maize Kernels Six Days after Silk Channel Inoculation with Fusarium graminearum
JOURNAL	Unpublished (2001)
COMMENT	Contact: Harris, Linda J. Eastern Cereal and Oilseed Research Centre Agriculture and Agri-food Canada Bldg. 21, Central Experimental Farm, Ottawa, Ontario, K1A 0C6, CANADA Tel: (613) 759-1314 Fax: (613) 759-6566 Email: harrislj@em.agr.ca.
FEATURES	Location/Qualifiers
source	1. .261 /organism="Zea mays" /cultivar="CO430" /db_xref="taxon:4577" /clone="Zm08_06h09" /clone_lib="Zm08_AAFC_ECORC_Fusarium_graminearum_inoculate d_corn_ear" /tissue_type="Developing kernels (sibcrossed)" /dev_stage="10-11 days post-silk emergence" /note="Vector: Bluescript SK+/XhoI-EcoRI; Site_1: EcoRI; Site_2: XhoI; Field-grown maize ears were silk channel-inoculated in the morning (~10 am) with 1 ml of a Fusarium graminearum macroconidial suspension (500,000 spores/ml) and whole ears were collected and immediately frozen in liquid nitrogen 6 days later."
BASE COUNT	65 a 60 c 59 g 61 t 16 others
ORIGIN	

Query Match 35.2%; Score 156.4; DB 11; Length 261;  
Best Local Similarity 78.6%; Pred. No. 1.7e-32;  
Matches 184; Conservative 12; Mismatches 34; Indels 4; Gaps 1;

RESULT 2  
 AI673919/c  
 LOCUS AI673919 129 bp mRNA EST 02-FEB-2000  
 DEFINITION 605039B05.x1 605 - Endosperm cDNA library from Schmidt lab Zea mays  
 cDNA, mRNA sequence.  
 ACCESSION AI673919  
 VERSION AI673919.1 GI:4874399  
 KEYWORDS EST.  
 SOURCE Zea mays.  
 ORGANISM Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
 clade; Panicoideae; Andropogoneae; Zea.  
 REFERENCE 1 (bases 1 to 129)  
 AUTHORS Walbot,V.  
 TITLE Maize ESTs from various cDNA libraries sequenced at Stanford  
 University  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Walbot V  
 Department of Biological Sciences  
 Stanford University  
 855 California Ave, Palo Alto, CA 94304, USA  
 Tel: 650 723 2227  
 Fax: 650 725 8221  
 Email: walbot@stanford.edu  
 Plate: 605039 row: B column: 05.  
 FEATURES Location/Qualifiers  
 source 1. .129  
 /organism="Zea mays"  
 /cultivar="Ohio43"  
 /db\_xref="taxon:4577"  
 /clone\_lib="605 - Endosperm cDNA library from Schmidt lab"  
 /tissue\_type="nucellar, embryo, and endosperm"  
 /dev\_stage="10-14 days post-pollination"  
 /lab\_host="DH5(alpha)"  
 /note="Organ: Kernel; Vector: pAD-GAL4-2'; Site\_1: EcoRI;  
 Site\_2: XhoI; Kernel endosperm cDNA library from Schmidt  
 lab"  
 BASE COUNT 32 a 26 c 25 g 46 t  
 ORIGIN

Query Match 27.3%; Score 121.2; DB 10; Length 129;  
 Best Local Similarity 95.3%; Pred. No. 6.1e-23;  
 Matches 123; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 316 atcatgaattcaagatactgCGGAGACATCATGATACTGCGGAGACAGACGGCCAGAGAT 375  
 ||||||||||| ||||||||||| ||||||||||| |||||||  
 Db 129 ATCATGAATTCATGATACTGCGGAGACATCATGATACTGCGGAGACAGACGGCGAGAGAT 70  
 Qy 376 gangctagctagatgccgtttcaccannatattatgtaacacccaaatctccatttttaa 435  
 || ||||||||||| ||||||| ||||||||||| ||||||||||| |||||||  
 Db 69 GAGGCTAGCTAGATGCTGTTTACCAAAAATATTATGTAACACCCAAATCTCCATTTTAA 10  
 Qy 436 gaaataaat 444  
 |||||||

Db 9 GAAATAAAT 1

RESULT 3

BG240000

LOCUS BG240000 309 bp mRNA EST 15-FEB-2001

DEFINITION OV1\_31\_G02.g1\_A002 Ovary 1 (OV1) Sorghum bicolor cDNA, mRNA sequence.

ACCESSION BG240000

VERSION BG240000.1 GI:12775073

KEYWORDS EST.

SOURCE sorghum.

ORGANISM Sorghum bicolor

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE 1 (bases 1 to 309)

AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt ,L.H.

TITLE An EST database from Sorghum: ovaries of varying immature stages

JOURNAL Unpublished (2000)

COMMENT Contact: Cordonnier-Pratt MM

Department of Botany

The University of Georgia

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860

Fax: 706 542 1805

Email: mmpratt@uga.edu

Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.

Seq primer: PolyTMix

High quality sequence stop: 305

POLYA=No.

FEATURES Location/Qualifiers

source

1. .309

/organism="Sorghum bicolor"

/db\_xref="taxon:4558"

/clone\_lib="Ovary 1 (OV1)"

/note="Organ: Mix of ovaries of varying immature stages

from 8-week-old plants; Vector: pBluescript II from Lambda

Zap II; Site\_1: XhoI; Site\_2: EcoRI; The library was made

from poly-A RNA in the cloning vector lambda ZAP II.

Clones to be sequenced were prepared by mass excision."

BASE COUNT 95 a 65 c 76 g 73 t

ORIGIN

Query Match 10.1%; Score 45; DB 11; Length 309;

Best Local Similarity 63.9%; Pred. No. 0.049;

Matches 85; Conservative 0; Mismatches 45; Indels 3; Gaps 1;

Qy 69 ggtgccttccaagaatcctccaaccacccttggtgcccaagcaagccacaaaaagtgggc 128

|| ||| | | || | || || | | |||| | ||||

Db 94 GGCTTCTTTCCATGGTTCTTCTGGCATCCTCAGTTGTTTATGCACGCACAATAAATGGGC 153

Qy 129 aaaccaaagaggacagcaatgctagggaaatgacgatgacaaagacgagggcatcgggca 188



```

Qy      69 ggtgccttccaagaatcctccaaccacccttggtgcccaagcaagccacaaaaagtgggc 128
      ||  ||| | | | | | | | | | | | | | | | | | | | | | | |
Db      94 GGCTTCTTTCATGGTTCCTCTGGCATCCTCAGTTGTTTATGCACGCACAATAAATGGGC 153

Qy     129 aaaccaaagaggacagcaatgctaggaaaatgacgatgacaaagacgagggcatcgggca 188
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     154 AAACCAAAGAGGACATCAACACCAGGAGTGTGACGATGAT--GACAAGGTCAGCAAGCT 210

Qy     189 acatacttggttag 201
      ||| ||| |||
Db     211 CCATAATTGGTAG 223

```

# ORIGIN

Query Match 10.0%; Score 44.6; DB 11; Length 1068;  
 Best Local Similarity 53.8%; Pred. No. 0.083;  
 Matches 92; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

```

Qy      1 cgaaaacactggtacccaaaacaaccgtcaaccaagggcaaattcaacaacctccaaaga 60
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     697 CGGACTCCCAGATACTGGAAACAACGCGGACCAAGAAACAAAGGAAAAGAAGGCAAACAG 756

Qy     61 ataatccgggtgccttccaagaatcctccaaccacccttggtgcccgaagcaagccacaaa 120
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     757 ACCACCAAGTACACACGCAGGAAAGCAGCAAAGACCGTAGCAGCCAAAGCAAGCCACAGA 816

Qy    121 aagtgggcaaaccaaagaggacagcaatgctaggaaaatgacgatgacaaa 171
      || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    817 AACAGGACACAAGAAACCAGGCACAAACGAGAAAAAAAAAAGGAGAACACA 867
  
```

RESULT 6  
 CNS015XR/c  
 LOCUS CNS015XR 1159 bp DNA GSS 26-JUL-1999  
 DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC  
 BACN15017 of DrosBAC library from Drosophila melanogaster (fruit  
 fly), genomic survey sequence.  
 ACCESSION AL106041  
 VERSION AL106041.1 GI:5619746  
 KEYWORDS GSS.  
 SOURCE fruit fly.  
 ORGANISM Plasmid Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 REFERENCE 1 (bases 1 to 1159)  
 AUTHORS Genoscope.  
 TITLE Direct Submission  
 JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :  
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
 - Web : www.genoscope.cns.fr)  
 COMMENT Determination of this BAC-end sequence was carried out as part of a  
 collaboration with the European Drosophila Genome Project (EDGP) -  
 http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC  
 library (Dros BAC) was made by Alain Billaud at CEPH (Centre  
 d'Etude du Polymorphisme Humain) with funding provided by a MRC  
 project grant. The DNA was prepared from embryos by Alain Bucheton  
 and Genevieve Payan. It has been constructed in the vector  
 pBeloBAC11.  
 FEATURES Location/Qualifiers  
 source 1..1159  
 /organism="Drosophila melanogaster"  
 /plasmid="pBeloBAC11"  
 /db\_xref="taxon:7227"  
 /clone\_lib="DrosBAC"  
 /clone="BACN15017"  
 /note="end : T7"  
 BASE COUNT 448 a 36 c 7 g 178 t 490 others



# ORIGIN

Query Match 9.2%; Score 41; DB 13; Length 1159;  
 Best Local Similarity 16.3%; Pred. No. 0.82;  
 Matches 53; Conservative 145; Mismatches 127; Indels 0; Gaps 0;

```

Qy      40 aaattcaacaacctccaaagaataatccgggtgccttccaagaatcctccaaccaccctt 99
      | | :|::|: :|::|: |: :::: : : : : : : : : | :
Db    1159 ASASASVAVVASASVAAVVSAAVSSVSSASASASSSSSSSSASSSSSMAAAAGVVS 1100

Qy     100 ggtgcccaagcaagccacaaaaagtgggcaaaccaaaggagacagcaatgctaggaaaat 159
      : : :|::|: :::::| : : : :|::| :|: | |::|: : |: ||:|
Db    1099 ASARSASAAVSVSAVSVSVVVASVSAVSVSAVVASSSASAAAARSAAVAVAVAAVAA 1040

Qy     160 gacgatgacaaagacgagggcatcgggcaacatacttggttagccgtaatgacgacgggcc 219
      :|: | |:| |:|: ::::| : : :|::| : |: | :| : : : : :
Db    1039 VMAAVMASASASASAVSVSSAVVSVASMASASVSASASCASAVAMSVVSVSSSAS 980

Qy     220 atgctatctagattccggtcttaatgagtacgtctgcagaaagactaataagtgtataa 279
      : : : : :| : : : : : : : : : : :| : : | | : : | |
Db     979 VSSSSSVSSSSASCRMSCAASAASAASVCGMSASMSMSAGASSVVSASAAASASASAASA 920

Qy     280 gagcttggtgctctgcgtggcgagttgtcaaccatcatcatgaattcaagatactgcgga 339
      :|: : : : : : : :|: : : : : :| : : : | : : : : : :|
Db     919 SAASASASASVASASASVSMASASMVVHASVVVVVASVSVSSASAVSMRVARVAAGVSA 860

Qy     340 gacatcatgatactgcggagacaga 364
      :|:| :|: :|:|:::|:|:|:|
Db     859 SASAAMABASWAVTVVSSASASASM 835
  
```

## RESULT 7

AW957338

LOCUS AW957338 674 bp mRNA EST 01-JUN-2000  
 DEFINITION EST369528 MAGE resequences, MAGE Homo sapiens cDNA, mRNA sequence.  
 ACCESSION AW957338  
 VERSION AW957338.1 GI:8147141  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 674)  
 AUTHORS Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt  
 ,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and  
 Quackenbush,J.  
 TITLE Assessment of gene expression patterns in a model of colon tumor  
 metastasis using a 19,200 element cDNA microarray  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: John Quackenbush  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 3528  
 Fax: 301 838 0208  
 Email: johnq@tigr.org  
 Plate: 106

Seq primer: Reverse.

FEATURES                      Location/Qualifiers

    source                    1. .674

                              /organism="Homo sapiens"

                              /db\_xref="taxon:9606"

                              /clone\_lib="MAGE resequences, MAGE"

                              /note="Vector: pBluescriptSKm"

BASE COUNT            233 a     133 c     158 g     150 t

ORIGIN

Query Match                      9.1%;    Score 40.6;    DB 10;    Length 674;

Best Local Similarity    54.3%;    Pred. No. 0.93;

Matches    82;    Conservative    0;    Mismatches    69;    Indels    0;    Gaps    0;

Qy     93    cacccttggtgcccaagcaagccacaaaaagtgggcaaaccaagaggacagcaatgcta 152

      || |||||    | | |||    |||||    | ||| |    ||    ||    |||||

Db     488   CAGCCTTGAAGCGAGGCCAAAAAGCAAAAGAAGTGCAGAAGAAGCTGGTGCATAATGCTC 547

Qy     153   ggaaaaatgacgatgacaaagacgaggcatcgggcaacataacttgttagccgtaatgacg 212

      |    ||    |    | | | |    |    |||| | | | |    |    || | |||| | | |||| |

Db     548   TGGCAAATTTGGAGTCTATGGGTAAAACATCAGGGAAGCTGTTTGATAGCAGTGATGATG 607

Qy     213   acggggccatgctatctagattccggtcttaa 243

      |||    |    | |    ||||    | |    |||

Db     608   ACGAATCTGATTCTTAAGATGACAGTAATAA 638

RESULT    8

CNS00L6W/c

LOCUS        CNS00L6W        1101 bp        DNA                      GSS                      03-JUN-1999

DEFINITION   Drosophila melanogaster genome survey sequence TET3 end of BAC: BACR24H20 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION    AL068145

VERSION      AL068145.1    GI:4958073

KEYWORDS     GSS.

SOURCE       fruit fly.

    ORGANISM   Drosophila melanogaster

                 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

                 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

                 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE    1 (bases 1 to 1101)

    AUTHORS    Genoscope.

    TITLE       Direct Submission

    JOURNAL     Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT      Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial

EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

```
FEATURES
    source
        Location/Qualifiers
            1. .1101
            /organism="Drosophila melanogaster"
            /db_xref="taxon:7227"
            /clone_lib="RPCI-98"
            /clone="BACR24H20"
            /note="end : TET3"
BASE COUNT      288 a      223 c      181 g      313 t      96 others
ORIGIN
```

Query Match 8.9%; Score 39.4; DB 13; Length 1101;  
 Best Local Similarity 32.2%; Pred. No. 2.2;  
 Matches 48; Conservative 47; Mismatches 54; Indels 0; Gaps 0;

```
Qy      15 cccaaaacaaccgtcaaccaagggcaaattcaacaacctccaaagaataatccgggtgcc 74
          ::::: :::::| :|:: |:|:: : |: ::::: :: | || || ||| |||
Db      1083 MVMVVVGSVSVGVGMAVSGCGVMRCRCMCMCMRMVMMCCAAAAAACCCGCGGGAC 1024

Qy      75 ttccaagaatcctccaaccacccttggtgcccaagcaagccacaaaaagtgggcaaacca 134
          |: :| | | :|| | |||| | ||||: :::: : :::: | :| : :
Db      1023 GGCMCMGCAGCAKCCCCCAACCCSCAAAGCCCAMCMRRVARRRRRRRGRRRGRRGRGGG 964

Qy      135 aagaggacagcaatgctaggaaaatgacg 163
          :| :| | :||| |:||| ||||
Db      963 RAARGAAGGSCAAGMAGCAGMAAACGACG 935
```

```
RESULT      9
AQ210844
LOCUS      AQ210844      771 bp      DNA      GSS      18-SEP-1998
DEFINITION HS_2230_A1_E06_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=2230 Col=11 Row=I, DNA sequence.
ACCESSION  AQ210844
VERSION    AQ210844.1 GI:3619813
KEYWORDS   GSS.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 771)
AUTHORS    Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
TITLE      Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
JOURNAL    Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE    99380589
COMMENT    Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
```

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Sequence Tagged Connector

Plate: 2230 row: I column: 11

Class: BAC ends

High quality sequence stop: 771.

FEATURES                      Location/Qualifiers  
    source                      1. .771  
                                /organism="Homo sapiens"  
                                /db\_xref="taxon:9606"  
                                /clone="Plate=2230 Col=11 Row=I"  
                                /clone\_lib="CIT Approved Human Genomic Sperm Library D"  
                                /sex="male"  
                                /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in  
                                E-Coli DH10B"  
BASE COUNT                      460 a           237 c           25 g           38 t           11 others  
ORIGIN

Query Match                      8.7%;   Score 38.6;   DB 13;   Length 771;  
Best Local Similarity           50.8%;   Pred. No. 3.4;  
Matches    92;   Conservative       0;   Mismatches    89;   Indels       0;   Gaps       0;

Qy       14   acccaaaacaaccgtcaaccaagggcaaattcaacaacctccaaagaataatccgggtgc   73  
         || ||| |||||   |||||   | || | ||||| |   | || |   |  
Db       585   ACACAACACAACACACAACCAAAACCCAAATAAACAAACCACCAACAACATAAACACCC   644  
  
Qy       74   cttccaagaatcctccaaccacccttggtgcccaagcaagccacaaaaagtgggcaaacc   133  
         |||   || ||   |||||   |   |||||   |   ||   || ||   |||  
Db       645   AAACCACAAAACCAACAACCAACTCAAAACCCCAATACACCCAAACAACAAAAAAAAT   704  
  
Qy       134   aaagaggacagcaatgctaggaaaatgacgatgacaaagacgagggcatcgggcaacata   193  
         ||| |   |||   ||   |   |   || |   || || ||   || |   || || |  
Db       705   AAACAATACAAAAAAACAACACCCACACCAACAAACACACAAAAACAACATACACCACA   764  
  
Qy       194   c   194  
         |  
Db       765   C   765

RESULT    10  
BG034748  
LOCUS       BG034748           1642 bp       mRNA                      EST           24-JAN-2001  
DEFINITION   602301702F1 NIH\_MGC\_87 Homo sapiens cDNA clone IMAGE:4403256 5',  
                                mRNA sequence.  
ACCESSION       BG034748  
VERSION       BG034748.1    GI:12428371  
KEYWORDS       EST.  
SOURCE       human.  
    ORGANISM    Homo sapiens  
                 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
                 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE       1   (bases 1 to 1642)  
AUTHORS       NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: DCTD/DTP  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM10113 row: c column: 01  
High quality sequence start: 18  
High quality sequence stop: 373.

FEATURES  
source Location/Qualifiers  
1. .1642  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4403256"  
/clone\_lib="NIH\_MGC\_87"  
/tissue\_type="mammary adenocarcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: breast; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.383 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH\_MGC Library."

BASE COUNT 534 a 471 c 389 g 247 t 1 others  
ORIGIN

Query Match 8.6%; Score 38.2; DB 11; Length 1642;  
Best Local Similarity 53.7%; Pred. No. 5.1;  
Matches 79; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

Qy 45 caacaacctccaaagaataatccgggtgccttccaagaatcctccaaccacccttggtgc 104  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1127 CGACGACCAACGAACAACACGCAGGAAGCACCACAACAACCAGACACGCACACACAAACA 1186

Qy 105 ccaagcaagccacaaaaagtgggcaaaccaaagaggacagcaatgctaggaaaatgacga 164  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1187 CAAAGCAACAAACACGAACACGACAGACGCAAGAAAAACACGAACGACAGAAGAACAGGA 1246

Qy 165 tgacaaagacgagggcatcgggcaaca 191  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1247 GAACAGAGACGAGAGCACCAAGCGACA 1273

RESULT 11  
BF671369  
LOCUS BF671369 697 bp mRNA EST 21-DEC-2000  
DEFINITION 602151249F1 NIH\_MGC\_81 Homo sapiens cDNA clone IMAGE:4292245 5', mRNA sequence.  
ACCESSION BF671369  
VERSION BF671369.1 GI:11945264  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens



```

CNS006ST
LOCUS      CNS006ST      937 bp      DNA      GSS      03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR14F16 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION  AL065880
VERSION    AL065880.1  GI:4944848
KEYWORDS   GSS.
SOURCE     fruit fly.
            ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE  1 (bases 1 to 937)
AUTHORS    Genoscope.
TITLE      Direct Submission
JOURNAL    Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT    Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.
FEATURES             Location/Qualifiers
     source            1. .937
                       /organism="Drosophila melanogaster"
                       /db_xref="taxon:7227"
                       /clone_lib="RPCI-98"
                       /clone="BACR14F16"
                       /note="end : TET3"
BASE COUNT      211 a      78 c      29 g      289 t      330 others
ORIGIN

```

Qy 135 aagaggacagcaatgctaggaaaatgacgatgacaaagacgagggcatcgggcaaca 191  
 ::: :| ::|: : ::||:| :| : |:| |:|::|:| :| :|: : :|  
 Db 714 GMRMMRARCRCMARASRMVVAAMAMAMCMRAAASAGASASRRGRGAACVVRGVGSA 770

RESULT 13

CNS0073W/c

LOCUS CNS0073W 922 bp DNA GSS 03-JUN-1999

DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR14D09 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL066784

VERSION AL066784.1 GI:4945247

KEYWORDS GSS.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 922)

AUTHORS Genoscope:

TITLE Direct Submission

JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES

source

Location/Qualifiers

1..922

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/clone\_lib="RPCI-98"

/clone="BACR14D09"

/note="end : TET3"

BASE COUNT 223 a 95 c 109 g 221 t 274 others

ORIGIN

Query Match 8.5%; Score 37.6; DB 13; Length 922;

Best Local Similarity 21.6%; Pred. No. 6.6;

Matches 37; Conservative 67; Mismatches 67; Indels 0; Gaps 0;

Qy 1 cgaaaacactggtacccaaaacaaccgtcaaccaagggcaaattcaacaacctccaaaga 60  
 | ::||: : :::: ::||:: : ::||::: :::: ::||::| :|::: |



```

Db      860  CAMMAAMNMMMACMMMMCMACMMAMCCMMACMMMAMAMMMMMMMMMAMMAMCACMAMMMA  801
Qy      61  ataatccgggtgccttccaagaatcctccaaccaccttggtgcccaagcaagccacaaa 120
      :  :      ::  ::  :  :| |||  ::  |      ::||  ::|  :||:|:
Db      800  CACMCAMMMCMMMMMMMMMCMCMCMCCACMMMACACMAMCCMMCMCMACMCMMAAM  741
Qy     121  aagtgggcaaaccaaagaggacagcaatgctaggaaaatgacgatgacaaa 171
      ||      ||::  ||:  :  |:|  ::|  :  :||:  :      |:|:|
Db     740  AAMMMMACAMMAMAAMMMMMMAMAAMMAAMMAAMMMMAAMMCCMCCMAAMAMA  690

```

# ORIGIN

Query Match 8.5%; Score 37.6; DB 13; Length 1128;  
 Best Local Similarity 33.3%; Pred. No. 6.9;  
 Matches 63; Conservative 44; Mismatches 82; Indels 0; Gaps 0;

```

Qy      3 aaaacactggtacccaaaacaaccgtcaaccaagggcaaattcaacaacctccaaagaat 62
      |:|:|:|: || | : :| | : | | |: | |:|: |:|:|:|:|: | |:|: |:
Db    1002 ARAAVASGGGVGGCGRGGMCAAASACGAAAAARCGAAAAAMMMMMAMMMMMCAMAMMM 943

Qy     63 aatccgggtgccttccaagaatcctccaaccacccttggtgcccagcaagccacaaaaa 122
      :: ::| | |:| :| || || || : | | ||| :| | : | ||:| |
Db     942 MMMAMRRRCGCGCCMCSVGAVAACCMAAAAAACASGCGMCCCACAAAMAAASAAAAAMAA 883

Qy    123 gtgggcaaaccaaagaggacagcaatgctaggaaaatgacgatgacaaagacgagggcat 182
      |:| |:| |:| |:| |:| |:| :: :: |:| |:| |:| |:| : | |:| :
Db     882 AGCNCAAVMACAVAAARAAAAAVACMVMMMSCAMCACAACMAAMACACMACSACRMACA 823

Qy    183 cgggcaaca 191
      || | |:|
Db     822 CGAACCAMV 814
  
```

## RESULT 15

CNS0075A

LOCUS CNS0075A 861 bp DNA GSS 03-JUN-1999

DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR14D11 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL066834

VERSION AL066834.1 GI:4945297

KEYWORDS GSS.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 861)

AUTHORS Genoscope.

TITLE Direct Submission

JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or

filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES  
source Location/Qualifiers  
1. .861  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone\_lib="RPCI-98"  
/clone="BACR14D11"  
/note="end : TET3"  
BASE COUNT 313 a 224 c 30 g 119 t 175 others  
ORIGIN

Query Match 8.4%; Score 37.2; DB 13; Length 861;  
Best Local Similarity 33.7%; Pred. No. 8.3;  
Matches 58; Conservative 40; Mismatches 74; Indels 0; Gaps 0;

Qy 3 aaaacactggtacccaaaacaaccgtcaaccaagggcaaattcaacaacctccaaagaat 62  
::|::| | | | | :|:| | | | :|:| | | | | | | :|:| | | | |  
Db 547 MMMAMMMACCCACCCACMAMAAACMMCMAMCAAAAAAAAAAAAAAMMMACAMAMAMCAAAM 606  
Qy 63 aatccgggtgccttccaagaatcctccaaccacccttggtgcccaagcaagccacaaaaa 122  
|: :: :: : | | | : | | : | | | :|:| :| | | : | | | | |  
Db 607 AMAMMAMCMMMMMCMAAAMAMACCMAMAACAMCMCAAAAAAAAAAMAMCMMHAAAAAAAAA 666  
Qy 123 gtgggcaaaccaaagaggacagcaatgctaggaaaatgacgatgacaaagac 174  
| | ::|: | : : |:: | | :| | : | | | :|:|  
Db 667 AAAAAACAAAMMAMAAMMMAMACMMAAAAAAAAAAMAAAMAMCAMCACCMCMCMC 718

Search completed: February 7, 2002, 08:20:45  
Job time: 18122 sec